



ARTICLE

Novel Rifle Number Recognition Based on Improved YOLO in Military Environment

Hyun Kwon^{1,*} and Sanghyun Lee²

¹Department of Artificial Intelligence and Data Science, Korea Military Academy, Seoul, Korea

²Graduate School of Information Security, Korea Advanced Institute of Science and Technology, Daejeon, Korea

*Corresponding Author: Hyun Kwon. Email: hkwon.cs@gmail.com

Received: 31 May 2023 Accepted: 25 July 2023 Published: 30 January 2024

ABSTRACT

Deep neural networks perform well in image recognition, object recognition, pattern analysis, and speech recognition. In military applications, deep neural networks can detect equipment and recognize objects. In military equipment, it is necessary to detect and recognize rifle management, which is an important piece of equipment, using deep neural networks. There have been no previous studies on the detection of real rifle numbers using real rifle image datasets. In this study, we propose a method for detecting and recognizing rifle numbers when rifle image data are insufficient. The proposed method was designed to improve the recognition rate of a specific dataset using data fusion and transfer learning methods. In the proposed method, real rifle images and existing digit images are fused as training data, and the final layer is transferred to the Yolov5 algorithm model. The detection and recognition performance of rifle numbers was improved and analyzed using rifle image and numerical datasets. We used actual rifle image data (K-2 rifle) and numeric image datasets, as an experimental environment. TensorFlow was used as the machine learning library. Experimental results show that the proposed method maintains 84.42% accuracy, 73.54% precision, 81.81% recall, and 77.46% F1-score in detecting and recognizing rifle numbers. The proposed method is effective in detecting rifle numbers.

KEYWORDS

Machine learning; deep neural network; rifle number recognition; detection

1 Introduction

Deep neural networks [1] perform well in image recognition [2–5] and speech recognition [6–9]. A deep neural network was applied to a closed-circuit television (CCTV) system [10–12] to enhance performance in real-time object recognition. A deep neural network can be linked to CCTV to recognize a specific person or object and provide the recognized information to the observer. In particular, there are many surveillance requirements in military situations; however, because the number of surveillance personnel is small, automated detection using CCTV is necessary. Especially, in a military environment, there is a need for a system that allows the model to detect objects by itself and provide warnings without a person by installing a deep neural network on such CCTVs. In military environments, object detection and recognition can be used to manage military equipment. There are



various types of equipment [13] in a military environment, and an automated management system is required. Among these pieces of equipment, the importance of managing the main firearm, the rifle, is high, and the firearm is currently managed by manually writing the rifle number. In this study, we propose a method for automatically detecting and recognizing rifle management.

In addition, the amount of data is insufficient because data collection is limited owing to security issues in military environments. Therefore, to recognize the rifle number, there is a need for methods to increase the rifle number recognition in a situation where the number of data is insufficient. Furthermore, we added a method to improve the detection performance of rifle numbers by combining data from rifle images published on the Internet with other types of numeric image datasets such as modified national institute of standards and technology database (MNIST) [14].

In this study, we propose a method for detecting rifle numbers using deep neural networks in situations where the amount of actual rifle image data is insufficient. The proposed method was designed to improve the recognition rate of a specific dataset using data fusion and transfer learning methods. In the proposed method, real rifle images and existing digit images are fused as training data, and the final layer is transferred to the Yolov5 algorithm model. This method detects the area corresponding to the rifle number and identifies the number of rifles in the detected area. In addition, a method for improving the performance of the proposed method was added by utilizing other datasets in situations where the amount of actual rifle data was insufficient. The contributions of this study are as follows. We propose a rifle number detection method that learns a real rifle dataset and another numerical dataset in a fusion format in situations where there is a lack of real rifle data. In terms of the experimental environment, we utilized the actual K-2 rifle dataset and fused various numerical datasets to conduct a comparative analysis of the performance.

The remainder of this paper is organized as follows. Related studies are explained in [Section 2](#). In [Section 3](#), the methodology is described. In [Section 4](#), the experiments and evaluations are explained. [Section 5](#) concludes the paper.

2 Related Work

This section describes deep neural networks, object recognition using deep learning, and data augmentation methods.

2.1 YOLO Model

The Yolo model is a deep learning model that detects objects by processing the entire image at once. Through this method, it is possible to detect the position of an object in an image and its classification. The Yolo model can be divided into object location detection and object classification.

In terms of object location detection, object position detection uses a rectangular bounding box to measure the length based on the coordinates of the four vertices of the detected object. The bounding box is a regression problem that detects an appropriate object location while reducing the difference between the correct value and the predicted value.

In terms of object classification, we use deep neural networks. The structure of a deep neural network comprises a combination of nodes and an input layer, a hidden layer, and an output layer. In the input layer, a value corresponding to the input data is assigned to each node, and the resulting value is activated through a linear operation such as multiplication and addition. If the value obtained through the activation function was less than a certain value, a value of 0 was provided; if it was more than a certain value, a value of 1 or the current value was provided. This varies depending

on the activation function, rectified linear activation unit (ReLU) [15], sigmoid [16], etc., and the calculated result value is transferred to the next layer. The hidden layer is internally composed of several layers, and as the number of layers increases, the number of computations increases. Prior to the improvement of computing technology, operation speed was slow; however, the development of computer technology has made operations faster through the use of parallel operations. In the hidden layer, if the value assigned to each node is less than a certain value through the activation function, the resultant value calculated through multiplication and addition is assigned a value of 0. In the output layer, the calculated value is transferred to the last hidden layer, and the probability value of each class is calculated according to the number of classes required through the softmax layer [17]. The largest value among the probability values was for the class recognized for the input data, and the sum of the probability values of each class was 1. In the deep neural network process, each node and weight parameter is optimized for the deep neural network using a cross-entropy loss function and a gradient descent method, while considering the correspondence between the predicted value for the input value and the actual class. Thus, after the deep neural network is completely trained, if new test data are provided as inputs to the network, a highly accurate prediction value is obtained. The proposed model, which is composed of a complex structure based on a yolo model, was used. The details are described in [Section 3](#).

2.2 Object Recognition Using Deep Learning

As recognition methods develop, recognition technologies based on specific colors and shapes, such as road signs and license plates, are being actively used. Methods for recognizing objects using deep learning include signboards [18–20], road signs [21–23], and vehicle license plates [24–26]. It is necessary to divide the boundaries of the signboard in the image to recognize signs of various colors and shapes. To clearly distinguish the boundary line of the signboard, a study is conducted to identify the boundary line of the signboard by photographing a signboard that emits light at night and searching only areas with high pixel brightness values. In the daytime environment, the recognition of a sign using an image is achieved by applying technology that recognizes a road sign or license plate.

Unlike signboards, road signs and license plates are produced to a fixed standard; therefore, they are of constant size. In addition, the color combination is simple so that people can easily recognize it. To recognize road signs, studies have been conducted to extract green and blue areas using the red, green, and blue (RGB) color model to detect the boundary of the sign and extract the boundary line, as well as methods to find the four vertices of the road sign using a polar coordinate system. There are various methods for recognizing license plates: recognizing the license plate of a vehicle based on a feature descriptor, recognizing a license plate pattern using a change in lighting, gray preprocessing, and a multistep process to find the arranged square size on the license plate. In this study, we developed a method for recognizing the numbers on rifles. Rifle numbers are smaller and less visible to humans than license plates, and the background and number of colors are almost similar. In addition, in a situation where it is limited to obtaining a dataset of rifle numbers owing to the military situation, a study was conducted on a number-recognition method that improved the recognition of rifle numbers.

2.3 Data Augmentation Methods and Data Fusion Methods

It is important to secure enough data using deep-learning technologies. However, in the healthcare field, biofield, and military fields, it is difficult to secure data owing to the privacy issues of patients and confidential military documents. When the image data are insufficient, they are augmented [27] using various linear techniques, such as rotating or expanding the original data and then cropping or flipping it up and down. In addition, methods for data augmentation using generative adversarial networks

(GANs) have been proposed [28]. However, these methods have a drawback in that the data cannot be formed into various distributions. In this study, the proposed method augments data using mirroring, random cropping, rotation, and shearing methods. Additionally, the proposed method uses data fusion after multiple datasets are built. We improved the recognition of rifle numbers by combining these datasets with other datasets. Even if the dataset was insufficient, a method for improving the number recognition of rifles in various distributions was applied by combining different types of datasets.

3 Proposed System

Data preprocessing is performed on the rifle number dataset used in the proposed method. As shown in Fig. 1, the rifle number-related dataset is first collected. After that, the rifle number dataset is scaled to a size of 640×640 . After that, mirroring, random cropping, rotation, and shearing are performed for data augmentation. After that, set the box bounding and each classification result value for the annotations, which are the correct values for each data.

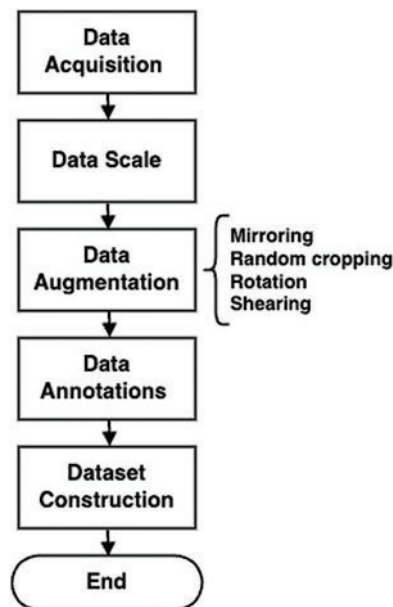


Figure 1: Overview of the pre-processing procedure

The proposed method is divided into a part that detects an object and draws a bounding box, and a part that detects a rifle number inside the bounding box. When detecting the rifle number in the bounding box, transfer learning is performed by applying a fusion method that mixes the real rifle dataset with other types of digit datasets. Transfer learning for the last layer trains the model to specialize in rifle number detection. In terms of the structural model, the proposed methodology is used to detect and recognize the number of rifles. This methodology uses a single neural network structure, predicts the bounding box, and shows the probability value of a recognized class for each test image. First, the model divides the entire image into multiple grids of specific sizes and generates the corresponding anchor boxes in each grid for the predicted scale and size. Each anchor box has an object score, x and y of the box center offset, box width, box height, and predicted class score. This model detects objects rapidly and enables one-stage object detection. In addition, it has good detection accuracy for objects.

In terms of the detailed structure, the methodology is a lightweight algorithm improved by Yolov3 and Yolov5 [29], and data augmentation, an activation function, and multi-graphics processing unit (multi-GPU) learning are performed. The proposed method consists of three parts: backbone, neck, and prediction, as shown in Fig. 2. The backbone extracts the features of each image. Because a cross-stage partial network (CSPNet) [30] has the advantage of fast processing time, CSPNet is used as a backbone and extracts feature maps from images. In addition, performance was improved by providing various input image sizes through a spatial pyramid pooling layer. The neck extracts feature maps in different stages. A path aggregation network (PANet) [31] was used as a neck to obtain feature pyramids. Feature pyramids help the model perform good detection on unseen data and generalize well in object scaling, which helps to identify the same object of different sizes and scales. The prediction classifies the input image and binds the object in the form of a box. In the prediction, anchor boxes are applied to each object and the final output vectors are generated using the class probability, object scores, and bounding boxes. Convolution, batch normalization, and leaky-ReLU (CBL) are composed of a combination of a convolution layer, batch normalization layer, and a leaky-ReLU activation function. CSP1 is composed of a CBL, residual unit (Res Unit), convolution layer, and concat layer. The residual unit is used in the residual structure and forms the basis of the CBL module. It also uses direct superposition of the tensor through the added layer. Res N contains n residual units of residual blocks and adds a CBL module and a zero-padding layer. CSP2 is composed of a CBL replaced with a residual unit from the existing CSP1. A detailed description of each component is described in Table 1.

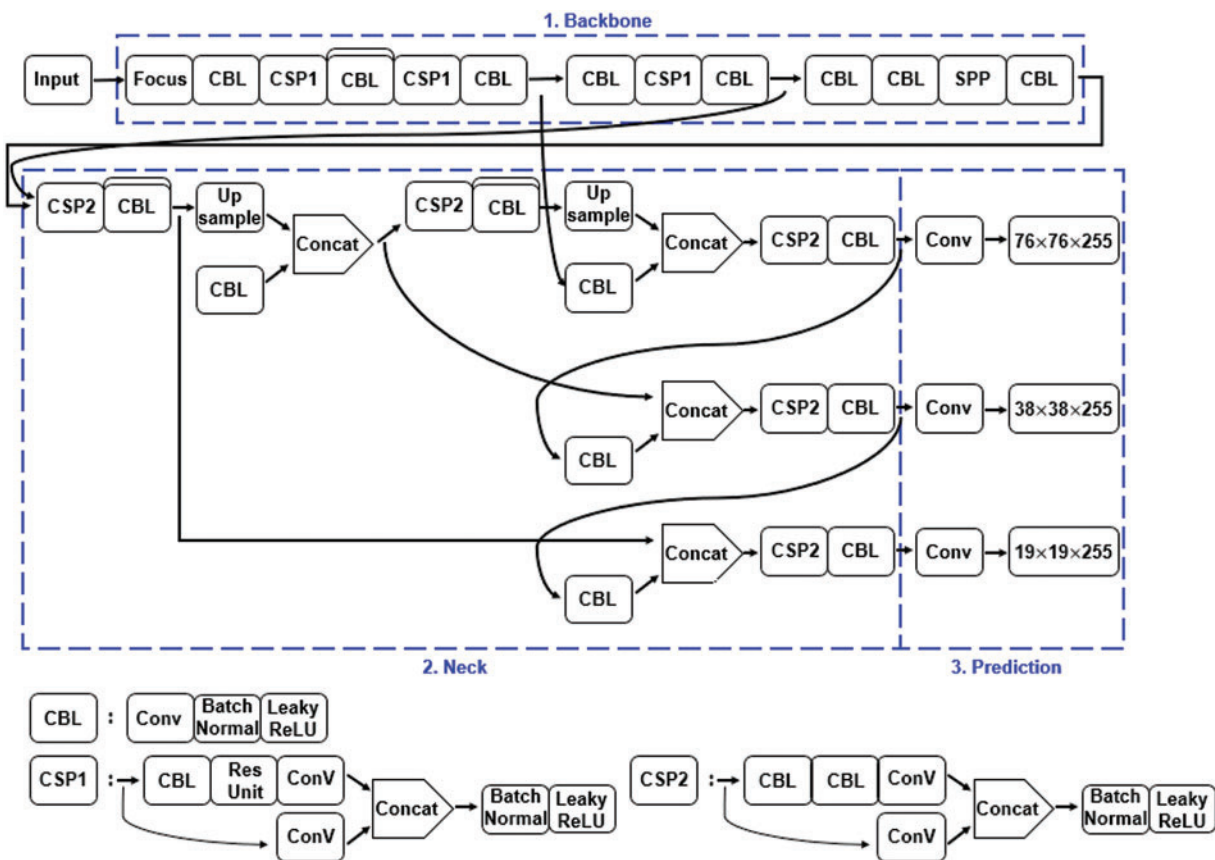


Figure 2: Overview of the methodology

Table 1: Characteristics table and components

Characteristics	Description
Cross-stage partial network (CSPNet or CSP)	It backbone and extracts feature maps from images. In addition, performance was improved by providing various input image sizes through a spatial pyramid pooling layer. The neck extracts feature maps in different stages.
Path aggregation network (PANet)	It used as a neck to obtain feature pyramids. Feature pyramids help the model perform good detection on unseen data and generalize well in object scaling, which helps to identify the same object of different sizes and scales. The prediction classifies the input image and binds the object in the form of a box. In the prediction, anchor boxes are applied to each object and the final output vectors are generated using the class probability, object scores, and bounding boxes.
Convolution, batch normalization, and leaky-ReLU (CBL)	Combination of a convolution layer, batch normalization layer, and a leaky-ReLU activation function.
Concatenate (Concat)	A method that connects an array in the direction of the selected axis.
Convolution layer (Conv)	Calculate as much data as the filter size as one value and use different filters as many as the number of output channels.
Spatial pyramid pooling (SPP)	A process of converting Region Proposals of different sizes into vectors of fixed size when passing a specific pooling process.
Batch normalization layer	Batch normalization is to place a normalization layer for each layer and adjust it so that the distorted distribution does not appear.
Residual block or residual unit	Passing the result of one layer not only to the next layer, but also to a later layer.
Leaky-ReLU	Set the output value to output a very small value such as not equal to when the input value is negative.
Upsampling	This is to increase the sampling period of the signal.

In terms of the rifle number-recognition procedure, the methodology is divided into a part that detects the area of the rifle number, and a method that recognizes each number in the detected area. When the input data were entered through the proposed model, the area corresponding to the number of rifles was first detected as a square, after which the process of recognizing each character through segmentation was performed for each character in the detected area.

4 Performance Evaluations

To demonstrate the performance of rifle number recognition, an experimental evaluation was performed using the proposed method implemented based on PyTorch, using images related to rifle numbers as a dataset. The firearm used in the experiment was a Korean K2 rifle.

4.1 Datasets

The dataset consisted of two main types. Because the number of actual rifle images was small, numerical image data were used. One was the rifle data with the rifle number visible in front and the other was the numerical image data. The dataset was built with 375 images, including actual rifle images and augmentation. The rifle number is located in the center, the rifle body part is taken as a standard, and is based on an image that is not inclined. The rifle number was labeled, the training data consisted of 300 pieces, and the test data consisted of 75 pieces.

The purpose of this dataset was to first detect the rifle number area in the rifle image, then cut out the area containing the rifle number and save it. Subsequently, the recognition accuracy of the rifle number can be increased by detecting only the rifle number in the area in which it is located. [Table 2](#) lists the training and test data for the four types. Another dataset, numeric pictures, was divided into four types to test the performance. In the first type: 1000 numbers in each font; in the second type: 25 images of rifle numbers and 1000 numbers in each font; in the third type: 2000 numbers for each font and pictures of 25 rifle numbers; in the fourth type: 949 MNIST images and images of 25 rifle figures.

Table 2: Training data and test data in four types

Description	Type 1	Type 2	Type 3	Type 4
Training data	800	820	1620	779
Test data	200	205	405	195

The numbers for each font are images written from 0 to 9 in 100 different fonts, and a dataset was constructed by inputting the numbers for each font using Microsoft Word and capturing and labeling the captured images. The numbers are white, and the background is black. A rifle number image is an image cut out by detecting only the rifle number part of the rifle image, and only the number part is labeled. In each dataset, the training data were divided into 80% and test data into 20%; the details are listed in [Table 2](#). The sizes of all images used in the training were designated as 416×416 . Separately, 11 images of firearms that were not used for training were prepared to test rifle number recognition.

4.2 Model Configuration

The detection and recognition models used a methodological model. The model was constructed using a binary cross-entropy loss function to calculate the loss of class probability and object score. Stochastic gradient descent (SGD) [32] was used as the optimization function. The model parameters

are listed in Table 3. The best experimental values were used as the numerical values of the parameters, listed in Table 3. There is a change in performance according to the change in each parameter; among them, the parameter value corresponding to the sweet spot is the result obtained experimentally.

Table 3: Model parameters

Parameter	Values
Learning rate	0.01
Momentum	0.937
Optimizer weight decay	0.0005
Warmup epochs	3.0
Warmup initial momentum	0.8
Warmup initial bias	0.1
Box loss gain	0.05
Cls loss gain	0.5
Anchor-multiple threshold	4.0
Image HSV-Hue augmentation (fraction)	0.015
Image HSV-Saturation augmentation (fraction)	0.7
Image HSV-Value augmentation (fraction)	0.4
Image flip left-right (probability)	0.5
Epochs	600
Batch size	24
Number of classes	80
Model depth multiple	0.33
Layer channel multiple	0.50

4.3 Experiment Results

Fig. 3 shows the process of detecting the area corresponding to the rifle number and recognizing each number in the detected area. The detection and recognition of the rifle number are performed, as shown in Figs. 3a–3d. First, we trained the model after assigning the labeling of the rifle number area in the image of the rifle. Consequently, it is possible to detect a location with a rifle number for an arbitrary rifle image, and only the corresponding area is cut using the bounding box x- and y-coordinates of the detected rifle number. The number of rifles was then determined for this cut-out area. Each identified rifle number is sorted in the order of the x-coordinates and then outputted as one continuous rifle number.

Fig. 4 shows the image data obtained using different types of numerical image datasets to increase the recognition of actual rifle numbers. The image was larger; however, for ease of identification, the image was cut such that the number part was visible, as shown in Fig. 4a. The dataset consisting of the rifle and augmented images shown in Fig. 4a was used to detect the location of the rifle number. Figs. 4b–4d show examples of the datasets used to recognize rifle numbers. Fig. 4b shows an example of the actual number of rifles. Fig. 4c shows an example of the font-specific numbers provided by Microsoft Word, and Fig. 4d shows an example from MNIST.

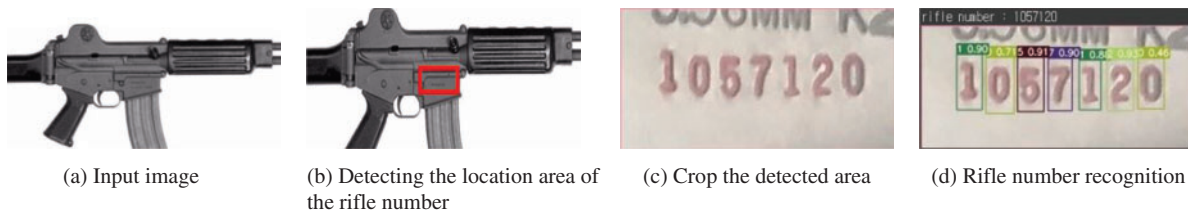
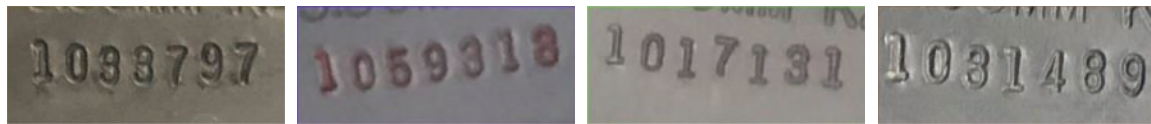


Figure 3: Process of detecting the area corresponding to the rifle number and recognizing each number in the detected area



(a) Image (the part number is clearly visible).



(b) Cropped image to label the detection area.



(c) Font-specific numbers provided in Microsoft Word.



(d) MNIST dataset

Figure 4: Image data using different types of numeric image datasets to increase recognition of the rifle number

Fig. 5 shows an example of the results of detecting the rifle number area. In the figure, the area corresponding to the rifle number in the model was detected well by the bounding box. Fig. 6 shows the results of recognizing the rifle number for each number in the cut-out rifle number bounding box. In the figure, each rifle number is recognized correctly with a high probability.

Table 4 lists the accuracy, precision, recall, and F1-score of rifle number recognition for each of the six cases. Accuracy is the percentage match between the number of rifles and the number recognized by the model. Experiments were conducted for six cases: Case 1:1000 numbers for each font; Case 2:1000 numbers for each font, 1000 license plates, and 25 rifle number images; Case 3:2000 numbers

for each font, 1000 license plates, and 25 rifle number images; Case 4:949 MNIST images and 25 rifle number images; Case 5: Case 4 with 1000 epochs; and Case 6:25 rifle number images.



Figure 5: Example of the result of detecting the rifle number area



Figure 6: Result of recognizing the rifle number for each number in the cut-out rifle number bounding box

Table 4: Accuracy, precision, recall, and F1-score of rifle number recognition for each of the six cases

Description	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Accuracy	0%	84.42%	75.32%	63.64%	61.04%	37.66%
Precision	0%	73.54%	73.74%	77.93%	72.93%	42.76%
Recall	0%	81.81%	75.90%	57.26%	55.60%	28.52%
F1-score	0%	77.46%	74.81%	66.02%	63.10%	34.22%
Number of correct answers per average rifle number	0	5.91	5.27	4.45	4.27	2.64

In terms of the detection performance for each case, Case 2 performed the best, with an accuracy of 84.42% compared to other cases. However, the case with the worst performance is Case 1, where the accuracy is 0%. Patient 1 did not detect any numbers in the picture of the rifle number. In Cases 1, 2, and 6, the position of the number could not be identified when there were only numbers in each font. However, in the case containing a labeled picture of the rifle number, the area of the rifle number was detected, but the exact number was not recognized in the detected area (average accuracy of 62.34%). However, in Case 2, which is a combination of Cases 1 and 6, the area of the rifle number was well detected, and the recognition of the rifle number in the detected area was good. Through this, the number location can be detected through the image of the actual firearm, and each number in the detected area can be correctly identified by learning a numeric dataset such as MNIST.

Cases 2–5 were cases in which a numeric dataset, such as MNIST, and a picture of a rifle number were used as the dataset. In Cases 2 and 3, the number of each font increased from 1000 to 2000, but

the rifle number recognition performance degraded owing to overfitting. In addition, in Cases 4 and 5, even though the number of training epochs increased from 600 to 1000, the detection performance was slightly lower. In Case 3, the MNIST image is human handwriting, and it is difficult to see that it has a certain shape compared to the number of each font; therefore, the actual rifle number is not recognized relatively well.

4.4 Experiment Analysis

Dataset. In the proposed method, we applied a recognition enhancement method by fusion of different datasets to improve rifle number detection. Therefore, MNIST, a numerical data set related to rifle numbers, and numerical fonts provided by Microsoft were built and trained as a dataset. To increase the recognition of rifle numbers, a numeric dataset such as MNIST, Microsoft Word fonts, and data augmentation were used as a data set. In addition, as shown in Table 4, the performance of the model was confirmed by combining various datasets for rifle number recognition in various ways from Case 1 to Case 6. It was possible to increase the detection rate of the rifle number area through the actual rifle number image, and it was possible to see that the recognition of the rifle number was high using MNIST and the font-specific numeric image dataset provided by Microsoft Word. In particular, because it has a characteristic pattern of rifle numbers and a number distribution similar to that provided by Microsoft Word, its performance is better than that of MNIST.

The process of detection and recognition. In this method, two processes were employed: detection and recognition. Using the proposed model, it can be seen that the rifle number area is first detected in the rifle image, and then the area detected is recognized by the rifle number. Therefore, the detection area must have good performance, and the detected area is separately stored and recognized by each number in the area. In addition, although the rifle number is a total of seven digits, the proposed method detects the number regardless of the number of digits in the rifle number, thereby enabling general number recognition.

Differences from license plate recognition. There is also a method that uses Py-Tesseract rather than the proposed method for numeric and letter detection, such as rifle number detection. However, it was confirmed through a separate experiment that Py-Tesseract is not suitable for detecting rifle numbers in images taken at various angles, lighting, and sizes and not in images captured by computer input text or scanned documents. Because Py-Tesseract has a limitation in meeting all the options required for image preprocessing (Gaussian blur, threshold, contour finding, contour range finding, etc.) to increase the detection rate of Py-Tesseract, its rifle number detection performance is poor. Therefore, instead of using Py-Tesseract, we used the proposed model to detect the rifle number area and recognize the rifle number.

The proposed method additionally trained on license plate dataset. The license plate dataset was configured and learned, as shown in Fig. 7. A license plate dataset consisting of 80,000 pieces of training data and 10,000 pieces of test data were used for training.



Figure 7: Image data using license plate datasets to train the proposed method

However, the rifle number detection performance of the proposed method was only 79.12%, resulting in performance degradation. This is because the license plate dataset differs from the rifle number dataset. First, the letters were different according to the angle difference, and the number of license plate data was included, so the type of letter class increased, the background color change according to the vehicle was reflected, and the number of license plate datasets was large; thus, the rifle number detection performance was rather poor.

When the amount of the actual rifle number dataset was comprehensively considered, the transfer learning method was judged to be effective. Transfer learning is a method of learning with a small amount of data by updating only the parameters of the fully connected layer, which is the last output layer, from a pre-trained model. The purpose of transfer learning is effective when the number of training data is small, and has the advantage of fast learning speed and high accuracy. In the case of license plate, there were differences as opposed to the actual rifle number data. In terms of the background, in the case of license plates, the background was black and white, green, or white. However, in the case of the rifle, the background was fixed to black and white. In terms of angle, the license plate dataset was taken from various CCTV images, and the angle was not constant, but in the case of the rifle dataset, there was some degree of consistency in angle because it was filmed through barcodes. In terms of the number of data, the number of license plate dataset was larger than the rifle number dataset, so the model was trained to be more suitable for license plate number detection rather than rifle number detection, which resulted in an imbalance in the rifle number detection classification performance.

Comparative analysis of accuracy and image processing speed for R-CNN, Faster R-CNN, and proposed method. In Table 5, we compared the proposed method and analyzed the regions using convolutional neural network (R-CNN) features and a Faster R-CNN model. R-CNN and Faster R-CNN are two-stage methods, whereas the proposed method is one-stage.

Table 5: Comparative analysis of accuracy and image processing speed for R-CNN, Faster R-CNN, and the proposed method

Description	R-CNN	Fatser R-CNN	Proposed
Accuracy	65.3%	70.4%	76.8%
Number of images processed per second	1	2	45

The R-CNN is a method for proposing a region using algorithms such as edge boxes and classifying within the region using a CNN. Faster R-CNN is a method that improves speed using a region proposal neural network (RPN) instead of edge boxes. In contrast, the proposed method simultaneously proposed the area and performs classification. Although its accuracy performance is slightly lower than that of the Faster R-CNN, it has the advantage of fast image processing per second.

5 Conclusion

In this study, we propose a method for recognizing rifle numbers using deep neural networks in situations where actual rifle image data are insufficient. The proposed method was designed to improve the recognition rate of a specific dataset using data fusion and transfer learning methods. In the proposed method, real rifle images and existing digit images were fused as training data, and the final layer was transferred to the Yolov5 algorithm model. In this method, the rifle number area was first detected in the input image and then the rifle number is recognized in the detected area. In

addition, a performance analysis was conducted using various data combination cases included in the numeric dataset. The experimental results showed that the proposed method correctly recognized the rifle number with 84.42% accuracy, 73.54% precision, 81.81% recall, and 77.46% F1-score. The proposed method can be used to recognize and manage rifle numbers in a military environment and can be used in conjunction with other methods.

In this study, rifle number detection has not been studied differently from previous studies, contributes to sparsity, and has the advantage of real-time rifle number detection. An interesting topic for future studies will be the development of a model that can classify rifle types in addition to detecting rifle numbers. In a situation where the number of rifles was not large, data were combined for each case to improve performance. The availability of military-related images is limited, but if there are sufficient images of rifles, performance can be improved. Although rifle number detection was improved using a unified model with the proposed methodology, the ensemble method could be the subject of future research. Additionally, the proposed method can be used to detect military equipment numbers, such as the serial numbers of communication equipment. And the applicability of the proposed method can be extended to address various security issues [33–36].

Acknowledgement: We thank the editor and anonymous reviewers who provided very helpful comments that improved this paper.

Funding Statement: This study was supported by the Future Strategy and Technology Research Institute (RN: 23-AI-04) of Korea Military Academy, the Hwarang-Dae Research Institute (RN: 2023B1015) of Korea Military Academy, and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A1A01040308).

Author Contributions: Study conception and design: H. Kwon and S. Lee; data collection: H. Kwon and S. Lee; analysis and interpretation of results: H. Kwon and S. Lee; draft manuscript preparation: H. Kwon. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data and materials used to support the findings of this study are available from the corresponding author upon request after acceptance.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Z. Zhan, J. Li and J. Zhang, “Evolutionary deep learning: A survey,” *Neurocomputing*, vol. 483, pp. 42–58, 2022.
- [2] Z. Mai, R. Li, J. Jeong, D. Quispe, H. Kim *et al.*, “Online continual learning in image classification: An empirical survey,” *Neurocomputing*, vol. 469, pp. 28–51, 2022.
- [3] H. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros *et al.*, “Transfer learning for medical image classification: A literature review,” *BMC Medical Imaging*, vol. 22, no. 1, no. 22, pp. 1–13, 2022.
- [4] Q. Zhang, Y. Xu, J. Zhang and D. Tao, “ViTAEv2 Vision transformer advanced by exploring inductive bias for image recognition and beyond,” *International Journal of Computer Vision*, vol. 131, pp. 1141–1162, 2023.
- [5] X. Ning, W. Tian, Z. Yu, W. Li, X. Bai *et al.*, “HCFNN: High-order coverage function neural network for image classification,” *Pattern Recognition*, vol. 131, pp. 1–11, 2022.

- [6] Y. Zhang, D. S. Park, W. Han, J. Qin, A. Gulati *et al.*, “BigSSL: Exploring the frontier of large-scale semi-supervised learning for automatic speech recognition,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1519–1532, 2022.
- [7] V. Bhardwaj, M. T. B. Othman, V. Kukreja, Y. Belkhier, M. Bajaj *et al.*, “Automatic speech recognition (ASR) systems for children: A systematic literature review,” *Applied Sciences*, vol. 12, no. 9, pp. 1–26, 2022.
- [8] P. Želasko, S. Feng, L. M. Velázquez, A. Abavisani, S. Bhati *et al.*, “Discovering phonetic inventories with crosslingual automatic speech recognition,” *Computer Speech & Language*, vol. 74, pp. 1–22, 2022.
- [9] P. Ma, S. Petridis and M. Pantic, “Visual speech recognition for multiple languages in the wild,” *Nature Machine Intelligence*, vol. 4, pp. 930–939, 2022.
- [10] P. Sukamto, I. Ispandi, A. S. Putra, N. Aisyah and R. Toufiq, “Forensic digital analysis for CCTV video recording,” *International Journal of Science, Technology & Management*, vol. 3, no. 1, pp. 284–291, 2022.
- [11] S. Ushasukhanya and M. Karthikeyan, “Automatic human detection using reinforced Faster-RCNN for electricity conservation system,” *Intelligent Automation & Soft Computing*, vol. 32, no. 2, pp. 1261–1275, 2022.
- [12] H. Kim, T. Kim and Y. Cha, “Methodology of displaying surveillance area of CCTV camera on the map for immediate response in border defense military system,” *Advances in Intelligent Systems and Computing*, vol. 1252, pp. 631–637, 2021.
- [13] C. Smith, K. Doma, B. Heilbronn and A. Leicht, “Effect of exercise training programs on physical fitness domains in military personnel: A systematic review and meta-analysis,” *Military Medicine*, vol. 187, no. 9, pp. 1065–1073, 2022.
- [14] D. Li, “The MNIST database of handwritten digit images for machine learning research,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
- [15] G. Vardi, O. Shamir and N. Srebro, “On margin maximization in linear and relu networks,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 37024–37036, 2022.
- [16] S. S. Atamanalp, “Endoscopic decompression of sigmoid volvulus: Review of 748 patients,” *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 32, no. 7, pp. 763–767, 2022.
- [17] J. Li, S. Lü and Z. Li, “Unsupervised domain adaptation via softmax-based prototype construction and adaptation,” *Information Sciences*, vol. 609, pp. 257–275, 2022.
- [18] S. Y. Arafat, N. Ashraf, M. J. Iqbal, I. Ahmad, S. Khan *et al.*, “Urdu signboard detection and recognition using deep learning,” *Multimedia Tools and Applications*, vol. 81, pp. 11965–11987, 2022.
- [19] J. Dai and Y. Gu, “OSO-YOLOv5: Automatic extraction method of store signboards in street view images based on multi-dimensional analysis,” *ISPRS International Journal of Geo-Information*, vol. 11, no. 9, pp. 1–20, 2022.
- [20] Y. Lin, W. Zeng, Y. Ye and H. Qu, “Saliency-aware color harmony models for outdoor signboard,” *Computers & Graphics*, vol. 105, pp. 25–35, 2022.
- [21] J. Wang, Y. Chen, Z. Dong and M. Gao, “Improved YOLOv5 network for real-time multi-scale traffic sign detection,” *Neural Computing and Applications*, vol. 35, no. 10, pp. 7853–7865, 2023.
- [22] Y. Zhu and W. Q. Yan, “Traffic sign recognition based on deep learning,” *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 17779–17791, 2022.
- [23] N. Dobrota, A. Stevanovic and N. Mitrovic, “Modifying signal retiming procedures and policies by utilizing high-fidelity modeling with medium-resolution traffic data,” *Transportation Research Record*, vol. 2676, no. 3, pp. 660–684, 2022.
- [24] H. Shi and D. Zhao, “License plate recognition system based on improved YOLOv5 and GRU,” *IEEE Access*, vol. 11, pp. 10429–10439, 2023.
- [25] H. Padmasiri, J. Shashirangana, D. Meedeniya, O. Rana and C. Perera, “Automated license plate recognition for resource-constrained environments,” *Sensors*, vol. 22, no. 4, pp. 1–29, 2022.
- [26] P. Kaur, Y. Kumar, S. Ahmed, A. Alhumam, R. Singla *et al.*, “Automatic license plate recognition system for vehicles using a CNN,” *Computers, Materials & Continua*, vol. 71, no. 1, pp. 35–50, 2022.
- [27] N. E. Nour, M. Loey and S. Mirjalili, “A comprehensive survey of recent trends in deep learning for digital images augmentation,” *Artificial Intelligence Review*, vol. 55, pp. 2351–2377, 2022.

- [28] X. Chen, Y. Li, L. Yao, E. Adeli, Y. Zhang *et al.*, “Generative adversarial U-Net for domain-free few-shot medical diagnosis,” *Pattern Recognition Letters*, vol. 157, pp. 112–118, 2022.
- [29] J. Li, C. Xu, H. Su, L. Gao and T. Wang, “Deep learning for sar ship detection: Past, present and future,” *Remote Sensing*, vol. 14, no. 11, pp. 1–44, 2022.
- [30] T. Mustaqim, C. Fatichah and N. Suciati, “Combination of cross stage partial network and ghostNet with spatial pyramid pooling on YOLOv4 for detection of acute lymphoblastic leukemia subtypes in multi-cell blood microscopic image,” *Scientific Journal of Informatics*, vol. 9, no. 2, pp. 139–148, 2022.
- [31] L. Zhou, X. Rao, Y. Li, X. Zuo, B. Qiao *et al.*, “A lightweight object detection method in aerial images based on dense feature fusion path aggregation network,” *ISPRS International Journal of Geo-Information*, vol. 11, no. 3, pp. 1–24, 2022.
- [32] Q. Li, D. Xiong and M. Shang, “Adjusted stochastic gradient descent for latent factor analysis,” *Information Sciences*, vol. 588, pp. 196–213, 2022.
- [33] J. Choi and X. Zhang, “Classifications of restricted web streaming contents based on convolutional neural network and long short-term memory (CNN-LSTM),” *Journal of Internet Services and Information Security*, vol. 12, no. 3, pp. 49–62, 2022.
- [34] Y. Lee and S. Woo, “Practical data acquisition and analysis method for automobile event data recorders forensics,” *Journal of Internet Services and Information Security*, vol. 12, no. 3, pp. 76–86, 2022.
- [35] J. Cabra, C. Parra, D. Mendez and L. Trujillo, “Mechanisms of authentication toward habitude pattern lock and ECG: An overview,” *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 13, no. 2, pp. 23–67, 2022.
- [36] N. Cassavia, L. Caviglione, M. Guarascio, G. Manco and M. Zuppelli, “Detection of steganographic threats targeting digital images in heterogeneous ecosystems through machine learning,” *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 13, no. 3, pp. 50–67, 2022.