International Journal of Internet, Broadcasting and Communication Vol.14 No.4 58-63 (2022)
http://dx.doi.org/10.7236/IJIBC.2022.14.4.58

IJIBC 22-4-8

A Study on Image Labeling Technique for Deep-Learning-Based Multinational Tanks Detection Model

Taehoon Kim, Dongkyun Lim*

Master's Student, School of Industrial and Systems Engineering, Georgia Institute of Technology, USA
**Professor, Department of Applied Software Engineering, Hanyang Cyber University, Korea*
*E-mail: tkim638@gatech.edu, *eiger07@hycu.ac.kr*

Abstract

Recently, the improvement of computational processing ability due to the rapid development of computing technology has greatly advanced the field of artificial intelligence, and research to apply it in various domains is active. In particular, in the national defense field, attention is paid to intelligent recognition among machine learning techniques, and efforts are being made to develop object identification and monitoring systems using artificial intelligence. To this end, various image processing technologies and object identification algorithms are applied to create a model that can identify friendly and enemy weapon systems and personnel in real-time. In this paper, we conducted image processing and object identification focused on tanks among various weapon systems. We initially conducted processing the tanks' image using a convolutional neural network, a deep learning technique. The feature map was examined and the important characteristics of the tanks crucial for learning were derived. Then, using YOLOv5 Network, a CNN-based object detection network, a model trained by labeling the entire tank and a model trained by labeling only the turret of the tank were created and the results were compared. The model and labeling technique we proposed in this paper can more accurately identify the type of tank and contribute to the intelligent recognition system to be developed in the future.

Keywords: *Convolutional Neural Network (CNN), Computer Vision, Deep Learning, YOLO Network*

1. Introduction

Among the various technologies of the 4th industrial revolution, artificial intelligence is by far the most interesting. The discussion of AI learning methods has become an important issue through the evolution from weak AI that simply enforces programmed rules to strong AI that solves problems through patterns obtained through self-learning. Among them, research on deep learning techniques that finds patterns on its own without humans providing a problem-solving method is active.

Problems that can be solved through deep learning techniques include visual intelligence, speech intelligence, and natural language processing. Among those problems, computer vision is a very active research field because it intuitively suggests solutions to many problems. In particular, efforts are being made to utilize it not only in industry but also in the field of public policy, especially in the field of national defense. For example, a model that can identify enemies in real-time by learning the corresponding image of the enemy personnel and weapon systems will lead to more efficient and accurate recognition of enemies and allies in wartime. It can also be used in various ways, such as using the model to fill the gap in closed-circuit television

Manuscript Received: August. 28, 2022 / Revised: September. 1, 2022 / Accepted: September. 5, 2022
Corresponding Author: eiger07@hycu.ac.kr
Tel: *** - **** - ****
Professor, Department of Applied Software Engineering, Hanyang Cyber University, Korea

footage monitoring of surveillance systems.

This paper discusses how to more effectively implement the visual intelligence of artificial intelligence mentioned above. In particular, it focuses on the importance of a labeling technique in the model development for object detection to specifically recognize the characteristics of an object to increase the efficacy of learning. The paper focuses to develop a model to identify tanks, the weapon system of the army in the field of defense, and a methodology for identifying tanks with higher accuracy through labeling suitable for the characteristics of the weapon system.

2. Related Theory and Prior Research

2.1 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a type of deep learning neural network mainly used to analyze visual images and videos. By applying the filtering technique to the artificial neural network, a convolutional operation is performed on the input structured in a grid format, and the operation to output a single value is performed in multiple layers. These layers are composed of convolution layers, pooling layers, and fully connected layers. The convolution layers and the pooling layers perform filtering, and then a classification operation is performed on the filtered image. The fully connected layers are the final output that is judged to be an accurate result. It also can be used up to the final classification by finding important features used for subsequent classification in the image by detecting the shape of an image through repeated convolution operation.[1] In this paper, we focused on the filtering technique of CNN and identified the essential features the neural network considers during the identification of an object called tanks.

2.2 You Only Look Once (YOLO) Network

YOLO network is one of the object recognition neural networks developed using the CNN technique described in 2.1. Through continuous improvement, the fifth version, YOLOv5, was recently developed. The YOLO network consists of 24 convolution layers and 2 fully connected layers, and a 448×448 image is input. A $7 \times 7 \times 30$ tensor is generated as a final output through convolution layers and pooling layers. The characteristic of YOLO network is that the operation speed is remarkably fast without a significant loss of accuracy. This was possible because the YOLO network calculates the probability for multiple bounding boxes and classes in a single neural network.[2] The YOLO network is deemed as a suitable implementation for the national defense field, which critically requires real-time object detection and identification.

2.3 Prior Research

Research to apply deep learning-based image processing techniques to the national defense field has been continuously conducted. There was a study that proposed a model for identifying multinational ships using a convolutional neural network.[3] Also, other group proposed a model that classifies helicopter types into binary in the same way.[4] Regarding the method to improve the performance of the model, recently, a technique was proposed to improve the recognition performance of the identification model, which is a camouflage group, using deep learning techniques.[5] Moreover, regarding tanks, a model was developed to detect and aim a tank by applying visual intelligence to image information[6], and other group developed a model to identify multinational tanks using the YOLO algorithm.[7] We also aims to improve the recognition performance of multinational tanks by proposing a new labeling technique for tanks.

3. Experiment and Result

3.1 Data Collection and Augmentation

Since the purpose of the model is to identify multinational tank objects, we set the class to 4 and collect data on tanks from 4 countries as shown in Table 1. Due to the nature of military equipment and its limitation to collect a large number of images, data augmentation techniques such as Flip, Gray, Hue, Noise, Crop, and Cutout were used to secure the learning data with the number of images below.

Table 1. Gathered images for classes

Class	Name of Tank (Country)	Number of Images (Augmented)
1	M1 (USA)	578 (518)
2	Merkava (Israel)	492 (437)
3	K1 (Republic of Korea)	534 (475)
4	T-72 (Russia)	532 (482)

3.2 Data Labeling according to Tank Features

Figure 1 shows a feature map with a CNN filter applied to each country's tank. The filtering results of an image shown on the left are shown in a sequential horizontal line with the 9th, 13th, and 17th layers respectively progressing to the right. The more critical the grid, the reddish the grid gets. To find out which part of the tank is featured by the CNN neural network, the feature map was calculated by randomly extracting 10 images for each class from the initially collected images. Here, the yellow box indicates the bounding part of the turret of the tank.

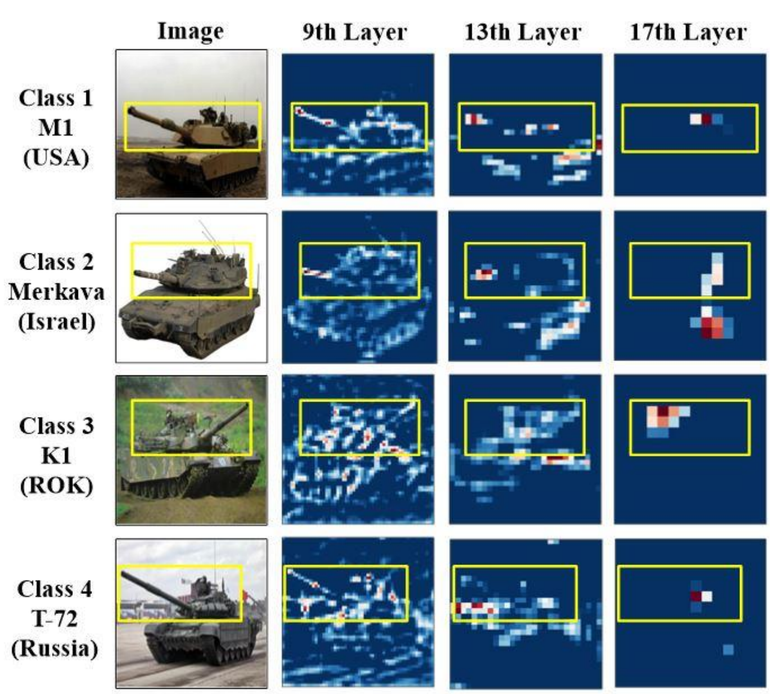

Figure 1. Example of CNN filtering to tank images

Moreover, the results of determining the number of images with the specific feature are shown in Table 2 below. It was found that, in general, CNN extracts turret as features from most tank images. If you label only the part that is mainly identified as a feature, you can create a model with less noise than the whole labeling and more effective performance. Therefore, when generating the object identification model in the future, the case of entire labeling and turret labeling were trained respectively to compare the model's performance.

Table 2. Result of CNN filtering

Heading level	Number of Random Images	Number of Turret Featured	Ratio
M1	10	9	90%
Merkava	10	6	60%
K1	10	10	100%
T-72	10	10	100%

3.3 Tank Identification Model Generation through YOLOv5 Network

The model generation to identify tanks from 4 countries was performed through the YOLOv5 network introduced in Chapter 2. As the development environment, Google Colab Platform was used, and learning was performed in the GPU environment. The ratio of the train image and the validation image was set to 8:2. Epoch was set to 200 times.

3.4 Model Operation Results and Analysis

As a result of the model operation, the targeted labeling of the turret improved the performance of the model compared to the general labeling of the entire tank with 1.5% accuracy, 5.4% recall, and 1.8% mAP improvement as shown in Table 3. It can be analyzed that the turret labeling in consideration of the features of the YOLOv5 network that mainly uses CNN as a filter, had a clear improvement. The graph of the learning process of the model is shown in Figure 2, and the graph is generated without overfitting. In addition, an example of the resulting object recognition for Validation Image is shown in Figure 3.

Table 3. Result comparison of each labelled model

Metrics	Entire Labelling	Turret Labelling	**Improvement**
Precision	94.9%	96.4%	**1.5%**
Recall	88%	93.4%	**5.8%**
mAP@ .5	93.9%	95.7%	**1.8%**

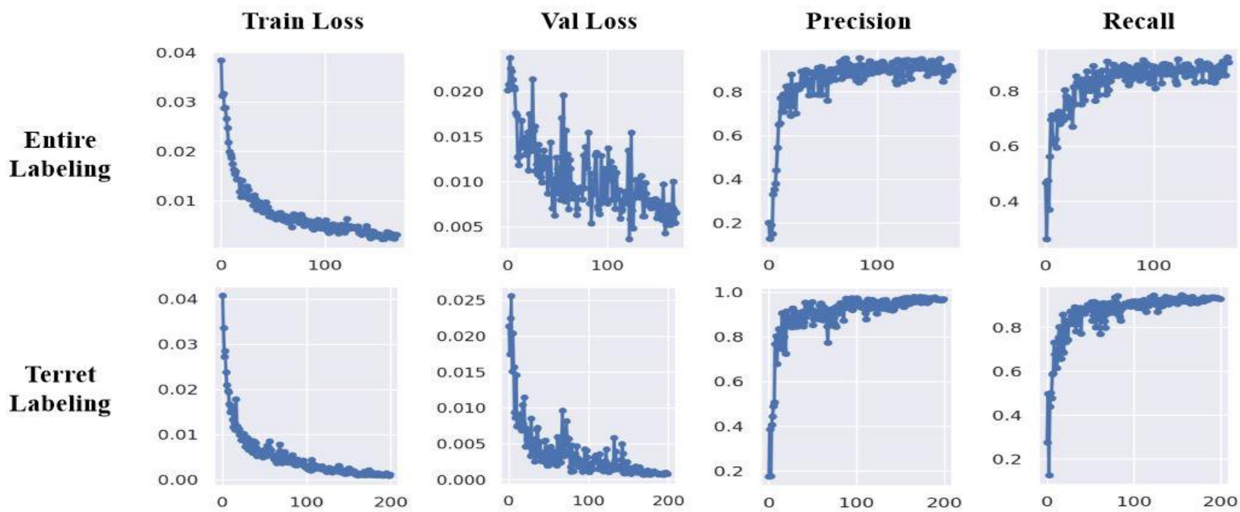

Figure 2. Learning progress of each model

Figure 3. Examples of each model's detection results

4. Conclusion

We presented a method to improve the performance of the model by applying a unique labeling technique when learning a deep learning object identification model for tanks. In terms of the combat technology concept of tanks, there are many cases where concealment and cover are primarily on the hull rather than the turret, or the hull is covered when maneuvering in the open field. Therefore, it is ideal to tactically set the aiming point between the turret and the hull which is the vulnerable point of tanks, and this paper can present an efficient way for the development of the aiming point calculation program for a tank using the bounding box of this model.

There are two limitations of this study. First, as there is limited open data on the topic of the defense field, the sufficient collection of data often limits the overall good performance in learning. Second, due to the nature of deep learning, it is difficult to identify the learning process and cause of the model. Therefore, pinpointing the correlation between the process difference in the labeling technique presented in this paper and the resulting recognition rate is not feasible.

For the future directions of the research, learning with a larger data set can be performed to further increase the recognition rate. Additionally, another research aspect could be to set different aiming points for other parts of tanks other than the turret to compare and, ultimately, increase the recognition rate. Alternatively, based on the identified tank turret, a model can be reevaluated to identify the appropriate aiming point.

References

[1] Matthew D. Zeiler and Rob Fergus, Visualizing and Understanding Convolutional Networks, ECCV 2014, Part I, LNCS 8689, pp. 818–833, 2014.
[2] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.
[3] J.-Y. Park, J.-H. Kim and H.-S. Moon, A Study on Expert System for Nationality Classification of Coast Guard Ship using Convolution Neural Network, Journal of Industrial Studies (J.I.S) 44, no.1: 45-60, 2020.
[4] T.-W. Kim, J.-H. Kim and H.-S. Moon, The Study on CNN based Helicopter Type Classification Model, Journal of Institute of Control, Robotics and Systems (J Inst Contr Robot Syst) 26, no.6: 479-486, 2020.
[5] G.-H. Choi, Performance Improvement of a Deep Learning-based Object Recognition using Imitated Red-green Color Blindness of Camouflaged Soldier Images, Journal of the Korea Institute of Military Science and Technology 23, no.2: 139-146, 2020.
[6] J.-H. Kim, C.-J. Jung, M.-R. Heo, Autonomous Battle Tank Detection and Aiming Point Search Using Imagery, Journal of the Korea Society for Simulation, vol. 27, no. 2, pp. 1–10, 2018.
[7] S.-G. Lim, D.-S. Kang, Identifications and Evaluation of Tank Nationality using YOLO Algorithm, KIISE Transactions on Computing Practices, 27(12), 555-562, 2021.
[8] S.-M. Jung, Watermarking Technique using Image Characteristics, International Journal of Internet, Broadcasting and Communication, 13.1: 187-193, 2021.
[9] M.-H. Song, Implementation of Low-cost Autonomous Car for Lane Recognition and Keeping based on Deep Neural Network model, International Journal of Internet, Broadcasting and Communication, 13.1: 210-218, 2021.
[10] H.-M. Lee and S.-J Lee, A Study on Security Event Detection in ESM Using Big Data and Deep Learning, International Journal of Internet, Broadcasting and Communication, 13.3: 42-49, 2021.
[11] S.-M. Jung, Advanced Pixel Value Prediction Algorithm using Edge Characteristics in Image, International Journal of Internet, Broadcasting and Communication, 12.1: 111-115, 2020.
[12] B.-S. Jang and S.-H. Lee, Sub-Frame Analysis-based Object Detection for Real-Time Video Surveillance, International Journal of Internet, Broadcasting and Communication, 11.4: 76-85, 2019.