International Journal of Advanced Culture Technology Vol.10 No.4 536-546 (2022)
DOI https://doi.org/10.17703/IJACT.2022.10.4.536

IJACT 22-12-63

Implementation of YOLOv5-based Forest Fire Smoke Monitoring Model with Increased Recognition of Unstructured Objects by Increasing Self-learning data

[1]Gun-wo Do, [2]Minyoung Kim, [3]Si-woong Jang*

[1]Undergraduate, Dept. of Computer Engineering, Dong-eui Univ., Republic of Korea
[2]Assistant Prof., Research Institute of ICT Fusion and Convergence, Dong-eui Univ., Republic of Korea
[3]Prof., Dept. of Computer Engineering, Dong-eui Univ., Republic of Korea*
dgw0601@naver.com, {kmyco, swjang}@deu.ac.kr

Abstract

A society will lose a lot of something in this field when the forest fire broke out. If a forest fire can be detected in advance, damage caused by the spread of forest fires can be prevented early. So, we studied how to detect forest fires using CCTV currently installed. In this paper, we present a deep learning-based model through efficient image data construction for monitoring forest fire smoke, which is unstructured data, based on the deep learning model YOLOv5. Through this study, we conducted a study to accurately detect forest fire smoke, one of the amorphous objects of various forms, in YOLOv5. In this paper, we introduce a method of self-learning by producing insufficient data on its own to increase accuracy for unstructured object recognition.

The method presented in this paper constructs a dataset with a fixed labelling position for images containing objects that can be extracted from the original image, through the original image and a model that learned from it. In addition, by training the deep learning model, the performance(mAP) was improved, and the errors occurred by detecting objects other than the learning object were reduced, compared to the model in which only the original image was learned.

Keywords: *Forest Fires, Unstructured Object, Deep Learning, YOLOv5*

1. INTRODUCTION

This year (2022), large forest fires broke out in several areas on the east coast of South Korea. To restore damage in this area, the Central Disaster and Safety Countermeasures Headquarters of the Republic of Korea invested a budget of 417 billion won. These fires cause a lot of social costs, and it also takes a long time to restore facilities and forests destroyed by it. If forest fires can be detected early on, the initial fire can be extinguished and damage can be reduced [1].

Local governments in the Republic of Korea have built a system related to this in order to detect forest fires at an early stage. Most of these systems are closed CCTV (Closed-circuit Television) systems, installing unmanned cameras in certain areas of the mountain and providing a way for the person in charge to directly monitor and visually check for forest fires at a remote-control center. Unfortunately, in most local governments, due to budget constraints, only a small number of people oversee monitoring forest fires via

Manuscript received: November 26, 2022 / revised: December 4, 2022 / accepted: December 9, 2022
Corresponding Author: swjang@deu.ac.kr
Tel: +82-51-890-2354
Professor, Department of Computer Engineering, Dong-eui University, Republic of Korea

video from multiple CCTV sources, making it virtually impossible to detect forest fires effectively.

There are systems that automatically monitor and report forest fires by introducing various technologies to supplement the problem mentioned above. First, there is an Internet of Thing (IoT) system-type monitoring system that detects smoke using various sensors. These sensors are installed and operated throughout the mountains. Compared to visually checking forest fires, this system has the advantage of being able to accurately figure out the location of forest fires. However, there is a problem of maintaining each installed sensor (battery replacement, checking whether it is installed normally), and each sensor could receive a false alarm due to mis-sensing that can occur in a large open space. Another method of monitoring forest fires is through satellites or unmanned aerial vehicles. Using such equipment has the advantage of expanding the scope of forest fire monitoring. However, in this case, monitoring circumstances largely depends on the condition of the atmosphere, and it is also difficult to monitor fires that start at a small scale. Finally, there is a system that monitors forest fires by installing additional thermal imaging cameras on existing CCTVs. It has the advantage of being able to quickly detect forest fires that are difficult to identify with simple CCTV images. However, this system also has the disadvantage of needing to be visually checked by humans [2-3].

In this paper, we introduce a study that implemented a deep learning model that can detect smoke from mountains to quickly detect forest fires even with existing CCTVs. The deep learning model in this paper was implemented using YOLOv5[4] and its performance was verified. In addition, research on self-amplifying data to increase the accuracy of monitoring (prediction) while holding a small amount of learning data is also introduced.

2. RELATED RESEARCH

2.1 Academic Research

Previously, computer vision fire detection technology was studied. The typical method would be detecting fire using the YCrCb color model studied[5]. Currently, with the development of deep learning technology, research on the implementation of a fire monitoring model using the YOLO[6] model is being actively conducted. The first related study was studied how to detect fire based on flames using YOLOv3[7]. The second related study was studied a method for fire inference on embedded equipment(Jetson Xavier NX) by learning photographs of blazes, flames, etc. using YOLOv4 and Tensor RT[8]. A third related study proposed a new atypical fire and smoke detection algorithm using deep learning and color histograms of fire and smoke[9]. Most of the above studies showed high accuracy because they were tested by giving images of clearly visible flames. If the fire detection method proposed in this study uses CCTV images installed at a high place or far away, it will only be possible to find it after the forest fire has spread to some extent.

The forest fires generate smoke before they occur. Therefore, we should study how to find smoke in order to detect forest fires early in CCTV images [3].

2.2 YOLOv5

There are various types of object detection models using deep learning. The object detection model needs speed and accuracy to be serviceable. However, accuracy(mAP) and detection speed (FPS) are in a trade-off, so when the detection speed is high the accuracy is low, and when the accuracy is low the detection speed is improved. Therefore, real-time detection requires a compromise between performance and model learning.

Among various deep learning object detection models, Yolov5 is a well-known model in the field of object detection. One-stage-detection methods such as Yolov5 have the advantage of being able to detect objects quickly.

The structure of Yolov5 consists of a backbone and a head. The backbone of Yolov5 is a part that extracts a feature map from an input image. The backbone of Yolov5 is implemented with CSPNet. CSPNet (Cross Stage Partial Network) has the advantage of achieving a lightweight model and effectively reducing memory cost while maintaining accuracy. As a result, CSPNet reduces the number of Bottleneck-CSP of YOLOv5 and helps lighten the model [10-11].

The head finds the position of the object based on the created feature map. First the anchor box is initially set, and afterwards the bounding box is finalized using the previous setting.

The Yolov5 models are in s(Fig.1 (a)), m(Fig.1 (b)), l(Fig.1 (c)), and x(Fig.1 (b)). For each model, the mean average precision (mAP) increases and the frame processing speed per hour decreases as it goes to the right in Figure 1. If we focus on object detection speed, we expect good results when we use YOLOv5s. Focusing on accuracy, Yolov5x can be seen as the most suitable model. However, according to Y. H. Jon (2021) [12], since the Yolov5l model is about 1-2% higher than other models, it is studied as the most suitable model as a trade-off between the accuracy of object detection and object detection speed.

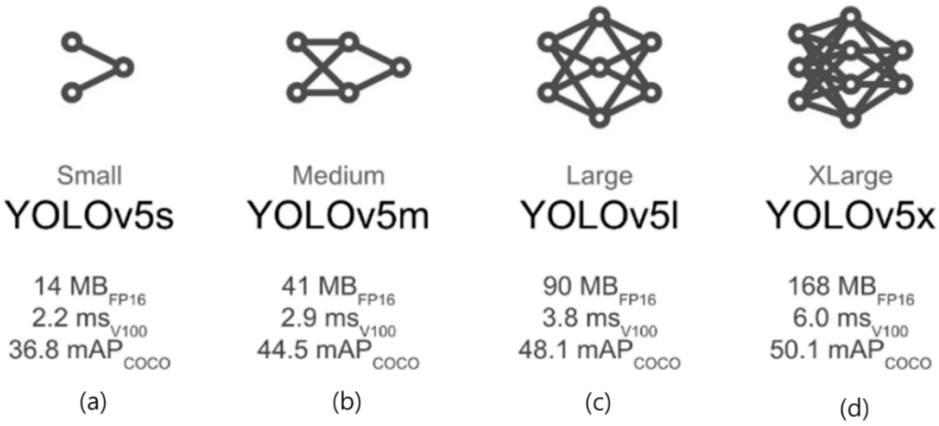

Figure 1. YOLOv5 Model

2.3 Object Detection Model Performance Evaluation Metrics

There are various performance evaluation indicators in deep (machine) learning. Among them, mAP, which is mainly used, is used when evaluating model performance of CNN (Convolutional Neural Network). In this case, the PR curve (precision-recall curve) and average prediction value (average precision) are methods used to evaluate the performance of object detection algorithms in computer vision. Although the evaluation may vary depending on how you apply the algorithm, it is not appropriate to evaluate the technology only with the detection rate, so both the detection rate and the accuracy are considered to properly evaluate the performance [13].

The PR curve is recognized as being detected when the threshold value exceeds a certain value according to the change of the threshold value for the confidence score (the probability that an object exists inside the bounding box when an object is detected). Since the precision and recall values change according to the change in the confidence score threshold, the PR curve is a graph that expresses this. The PR curve is good for understanding the overall performance of an algorithm, but it is inconvenient to quantitatively compare the performance of different algorithms, so the concept of average precision came out. The higher the average precision, the better the performance of the algorithm, and mAP (mean average precision) is obtained by calculating the area of the graph from the average precision graph and dividing each area by the number of object classes.

3. A MODEL FOR WILDFIRE SMOKE DETECTION

This paper studied the process for creating a Yolo model for designing a forest fire smoke monitoring system using the Yolov5-based Yolov5l model.

Since the smoke object in the forest fire image is an amorphous object, it has various forms. Although the performance of the learning model improves when learning from datasets that has its object classes segmented and unnecessary data removed, it has not been verified whether the performance of deep learning is guaranteed when the same object is segmented by shape [14]. In addition, when learning smoke objects that exist within a specific space (bounding box) to secure forest fire images, the source data are natural images that are likely to have various objects in the bounding box, so object detection can be difficult [15]. To overcome this problematic situation, the corresponding method was sought in this paper.

3.1 Configuring Datasets

The dataset for the model to learn was obtained from Roboflow[16]. Roboflow collects forest fire smoke data and extracts the characteristics of objects to be learned, so we collected 3,707 images that can be learned well, have clear image quality, and focus on smoke from forest fires. The test data also went through the same process and similarity with the training data was minimized. Forest fire smoke objects have an atypical shape, but learning occurs through a deep learning model when objects, and the situation in which the objects occur are embodied as smoke from forest fires.

3.2 Extracting Objects

3,707 pieces of collected data are insufficient for training a deep learning model. Therefore, in this paper, we present a self-increasing method to learn efficiently from small amounts of data.

Deep learning models are weighted through the process of displaying bounding boxes that display objects in images containing objects to learn objects, and by learning coordinates pointing to them. Learning data plays an important role in the learning process of deep learning models, but it is difficult to collect and refine unstructured data. Therefore, it was confirmed that if the dataset, created through the process of extracting the crop image from the already collected and refined original dataset and adding it back to the original dataset, is performed with the same learning parameters, it shows a higher mAP value than the learned weight of the original dataset.

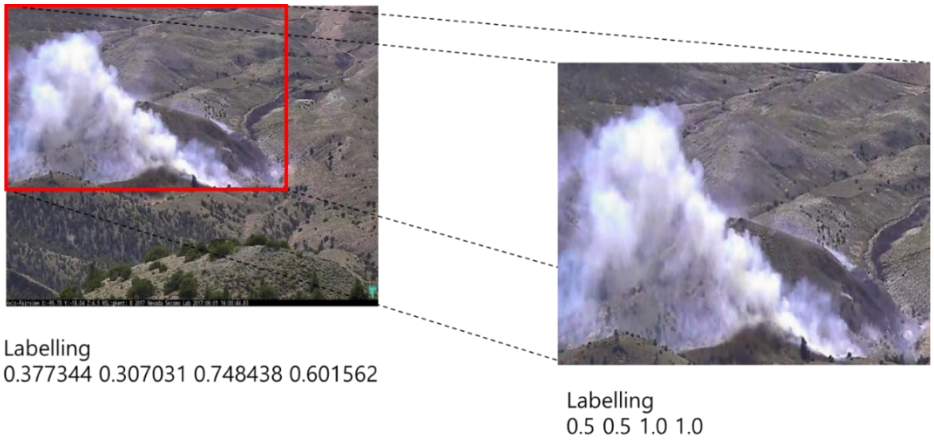

Figure 2. Example of crop image extraction

The left image of Figure 2 is image data learned by the labeled coordinates of the original image and the original dataset weight. The right image is re-learned by labeling the entire image on the Crop image extracted from the original through the learned weight, combining the extracted image with the original image dataset, resulting in the weight in Figure 2.

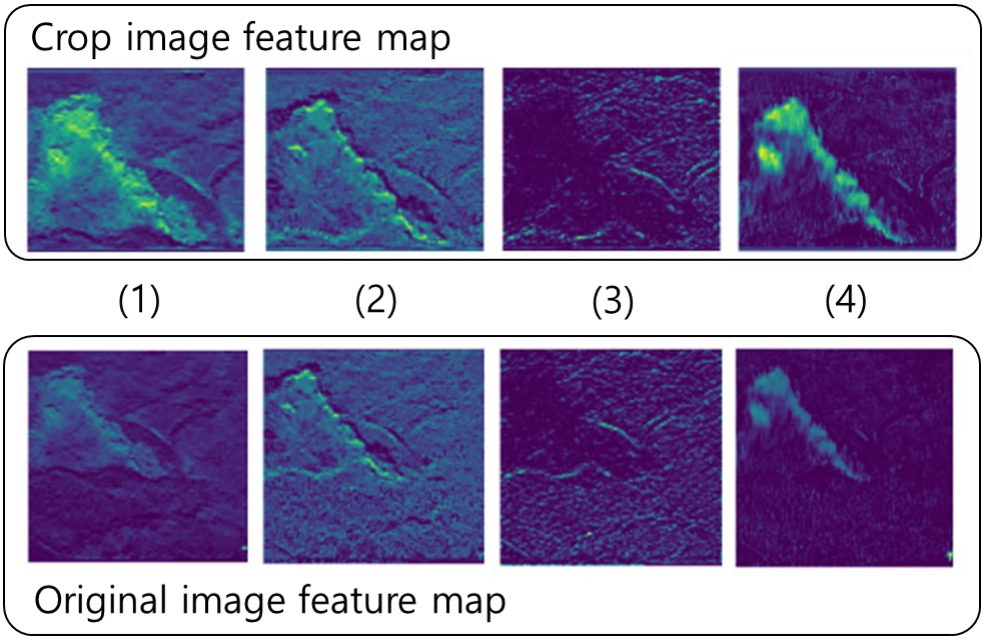

Figure 3. Compare feature maps of each image

When viewed from a human perspective, the left image and the right image of Figure 2 are judged as the same object in the same situation. When computer vision, such as a learned machine learning model, analyzes image patterns, there is a difference in correlation between adjacent pixels of feature maps that extract features of objects, given the feature map in Figure 3. Figure 3 is an image that shows the feature map of Step 4 out of the 24 feature maps of the image extracted from the image in Figure 4 with the weight with the mAP value in Figure 3.

Each feature map has a rougher pixel and deeper channel depth from left to right, as it passes through the deep neural network. In the deepening process, CNN neural networks find the features of the image through the convolution layer and the pooling layer, then adjust the weights and computations required for image processing. It is found that the feature extraction of objects from the convolution layer of the first and fourth images from the left of Figure 3 is more active in the additional learned weight of the original crop image than in the weight of the original-learned dataset.

3.3 Multi-time Object Detection Feature Map Comparison

We determined how many changes occur in the Feature Map in the process of detecting an object, extracting it from the original image by the size of the Bounding Box, and then detecting it with the same model again.

During the process of continuously extracting and detecting objects, when it came to the point that the size of the bounding box decreased and did not change in certain rounds, we compared the feature maps in the same convolution layer for each turn.

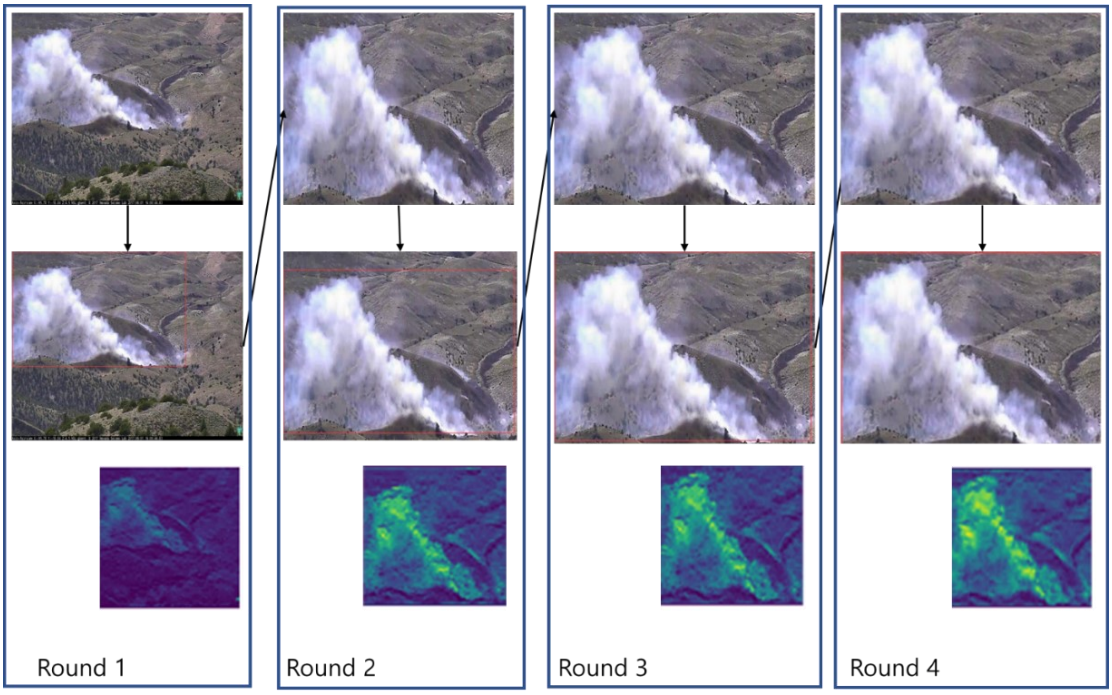

Figure 4. Feature Map of Objects by Round

The arrow from the top to the bottom of Figure 4 means the detection of an object, and the arrow from the bottom left to the top right means the extraction of the object. The process of detecting an object from an image corresponds to the first round. The model used to extract the corresponding feature map was weighted showing the mAP performance in Figure 4. The model used for feature map extraction learned the images in Round 1, and a fixed labeling of {0.5 0.5 1.0} on images extracted from images in Round 1.

The image below Figure 4 compares a. the same image for each round, with b. the crop image extracted from image a., and c. the image of stage0_Conv_features among the feature maps extracted also from image a.The feature map of the first and second rounds showed a significant change in the degree of activation, but from the third round and onward, there was no significant change from the feature map of the second round.

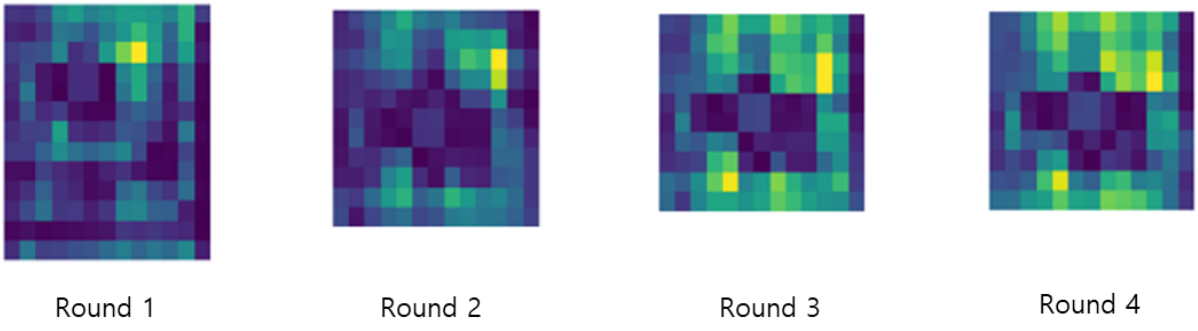

Figure 5. YFeature Map result Comparison of YOLOv5 Deep Neural Networks

What can be seen from Figure 5 is that there was no significant change from the third round in the shallow neural network stage0_conv_feature. However, images of stage_C3_features, a deep neural network, show the activation of an additional feature map when comparing the second and third rounds.

4. IMPLEMENTATION RESULT

In this paper, as an atypical feature of the wildfire smoke object of the bounding box, background noise, other than the wildfire smoke object, of the image was learned. Learning of images containing backgrounds in real environments showed false positives in actual detection results. However, images as large as the bounding box are extracted from images through detection with the weight of the original learning model, a label of {0.5 0.5 1.0 1.0} is generated on the extracted image, and the sum is added to the image dataset, and the trained model and the extracted image are generated. The performance improvement compared to the original learning model is shown in Figures 6~8.

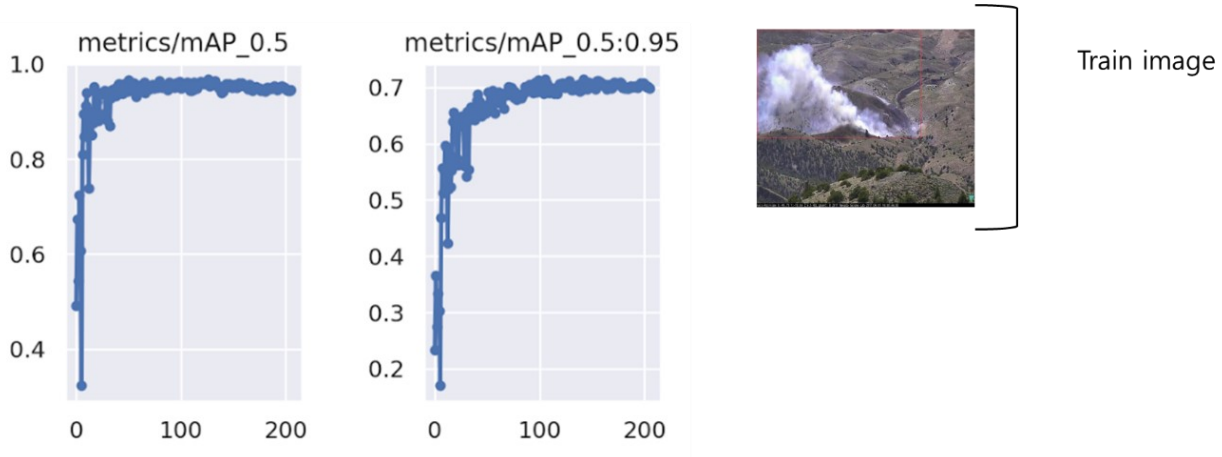

Figure 6. The mAP result of the original learning model (Model Case 1)

Figure 6 shows the mAP result of the model trained on the original dataset and the train image constituting the training dataset.

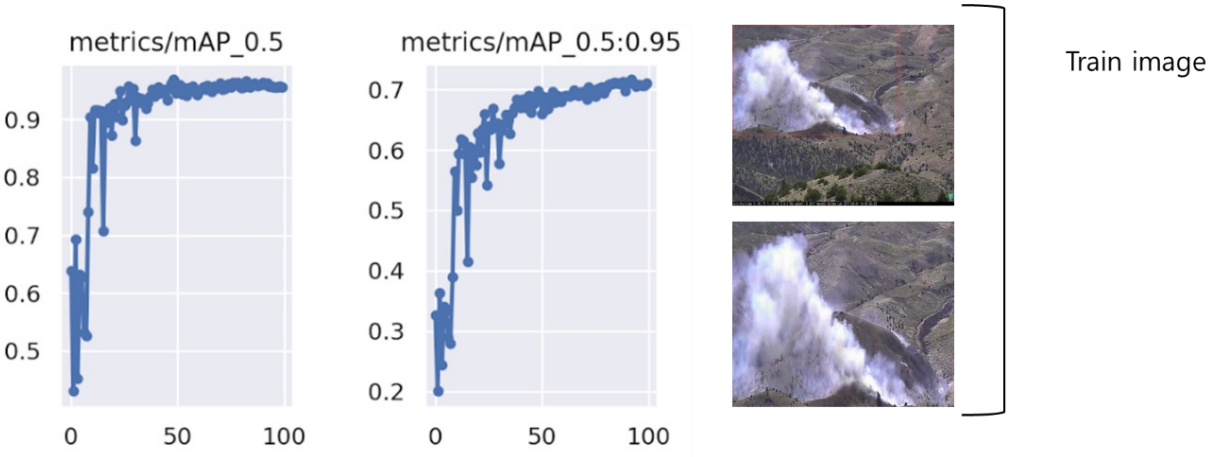

Figure 7. The mAP result of the original image and the image training model extracted from the original (Model Case 2)

Figure 7 shows the mAP of the model trained with the original image and the crop image extracted from the train image in Figure 8.

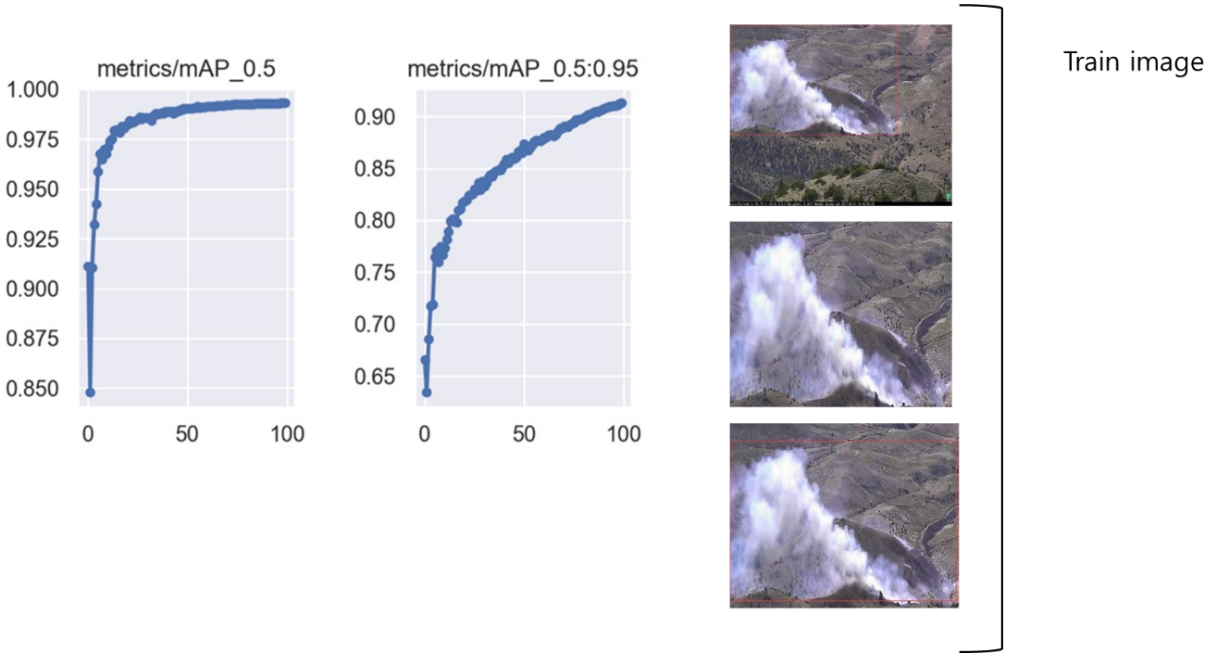

Figure 8. The mAP result of the model that learned the labeling of the original image and the original image, along with the image extracted from the original (Model Case 3)

Figure 8 shows the mAP of the result of training with the training dataset in Figure 7, and the extracted crop image in Figure 7 added with labeling detected by the model in Figure 6.

Figures 6 through 8 compare the learning results of the three learning model weights above under the assumption that an effective dataset can be created by extracting images up to 3 rounds when the correlation of each pixel is checked in the shallow neural network and the deep neural network.

Table 1. Deep learning model learning environment

OS	CPU	RAM	GPU
Windows 10	I5-10400	16GB	GeForce RTX 3080Ti (GDDR6X 12GB)

Table 2. Learning parameters and performance results for each model case

Case	Train			Validation	Epochs	batch size	YOLO Model	Performance value	
	original	crop	crop detect					mAP50	mAP50-95
1	3,378			376	100	16	YOLOv5l	0.953	0.712
2		3,308						0.966	0.716
3			1,506					0.958	0.718

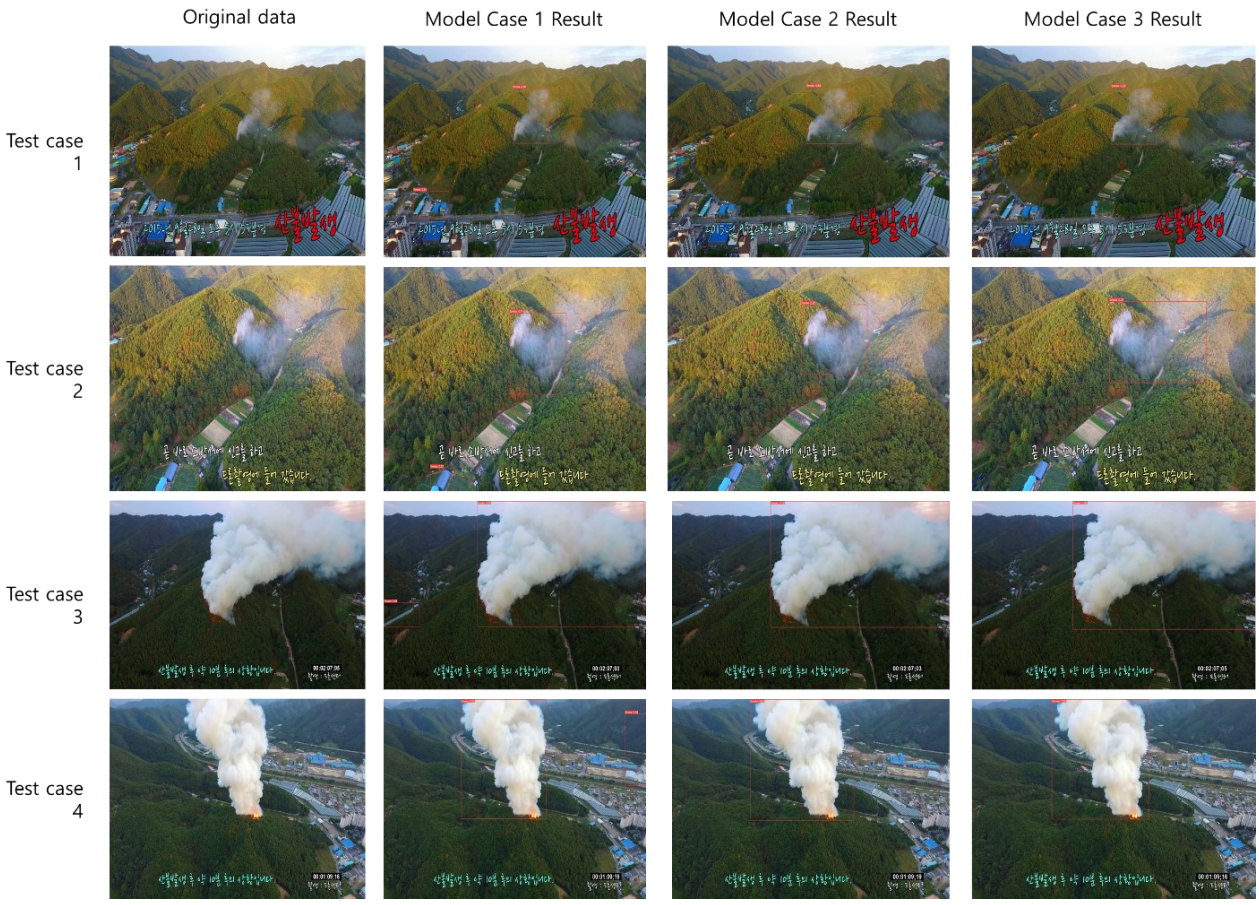

Figure 9. Actual Wildfire Smoke Detection Results for Each Model Case

The original data in Figure 9 are images found from aerial videos of actual forest fires, which were not included in the learning and testing data throughout the Model Cases (Figure 6~8). While the deep learning model learns, it learns not only the objects but also the noise data, and we can verify these mis-detected objects in Test Case 1 of Figure 9. But in Test Case 2 and 3 of Figure 9, we can see that no mis-detected objects appear.

Table 1 shows the computer environment used to learn the model, and Table 2 shows the learning parameters and performance results used in Model Case 1 (Figure 8), Model Case 2 (Figure 9), and Model Case 3 (Figure 10), respectively.

Figure 9 shows the actual detection results of Model Cases 1, 2, and 3. Looking at the test results, Case 1 Result is the result of detection with the weighted model of Case 1 in the original data. The result was that 1 smoke object and 1 unrelated object were falsely detected. However, comparing the detection results of each model in Case 2 and 3, it can be confirmed that the weighted model in Case 3 detects smoke objects and does not detect false positive objects in Case 2. However, comparing Case 2 and Case 3, there is only a slight difference in the size of the bounding box, and no significant difference exists. When considering the feature map change between Round 1 and Round 2 in Figure 6, the degree of improvement in mAP performance for each case, and the amount of time required to build the dataset, it can be seen that Case 2 is the most effective.

5. CONCLUSION

In this paper, based on the deep learning model YOLOv5, we present a deep learning method through efficient image data construction to monitor forest fire smoke, which is unstructured data.

Wildfires are characterized by smoke being detected first. Acting is an unstructured object and is not suitable for the learning method of existing object detection models. The boundary between the learning object and the background inside the bounding box is not clear. Due to these characteristics, even if the learned model shows a high mAP value, many misdetections occur in real-world tests.

In this paper, it is confirmed that the activity of the feature map is added when learning by extracting objects by the size of the bounding box from the original image and including them in the learning dataset. Although continuous data addition to the original dataset is possible by repeating the object separation and rediscovery process, it showed effective performance improvement in the first object separation process.

When a dataset is constructed by the method of this paper, it is possible to reduce the cost of building the dataset because it is extracted from the original image, and it shows a 1.3% mAP increase compared to the learning dataset of this paper. In addition, it was confirmed that the false detection case was reduced compared to the existing model in the actual test image.

Currently, a large amount of learning data is needed to provide high-quality artificial intelligence services. However, conventional learning data collection and purification methods require a lot of work time and money, making it difficult to secure enough learning data within a limited budget and physical time. However, we show here that the original image and the original image learned model can improve the performance and reduce the case of mis-detecting when we construct a dataset with a fixed label {0.5 0.5 1.0 1.0} and let the deep learning model learn from it. It is expected that the collection method presented in this paper can be applied to secure efficient learning data.

ACKNOWLEDGEMENT

This Work was supported by Dong-eui University Foundation Grant (2021). Also, This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2022-2020-0-01791) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation). Finally, Thank you to Donggyu Choi (dgchoi @deu.ac.kr) senior researcher who helped me a lot when I wrote this paper.

REFERENCES

[1] M. N. Jang and Y. J. Lee, "Selecting Forest Fire Vulnerable Regions Based on Biophysical and Socioeconomic Factors in Gangwon Province," *Crisisonomy (KRCEM)*, Vol. 16, No. 3, pp.133-144, 2022. DOI : http://doi.org/10.14251/crisisonomy.2020.16.3.133

[2] S. J. Oh, "Database Design for Growth Prediction of Forest using Drone Photo," *The Journal of the Convergence on Culture Technology (JCCT)*, Vol. 6, No. 4, pp.709-714, 2020. DOI: http://koreascience.or.kr/article/JAKO202034965718305.page

[3] B. C. Ko, "IoT technology for wildfire disaster monitoring," *Broadcasting and Media Magazine*, Vol. 20, No. 3, pp.91-98, 2015. DOI: https://koreascience.kr/article/JAKO201524848982493.page

[4] YOLOv5, https://github.com/ultralytics/yolov5

[5] S. Lee, B. Shin, B. Song, S. Song, S. An, J. Kim and H. Lee, "Wild Fire Monitoring System using the Image Matching," *The Journal of the Korea Contents Association*, Vol.13, No. 6, pp.40-47, 2013. DOI: https://doi.org/10.5392/JKCA.2013.13.06.040

[6] YOLO: Real-Time Object Detection, https://pjreddie.com/darknet/yolo/

[7] Y. Kim and H. Cho, "Detecting Location of Fire in Video Stream Environment using Deep Learning," *The Transactions of the Korean Institute of Electrical Engineers*, Vol.69, No. 3, pp.474-479, 2020. DOI: http://doi.org/10.5370/KIEE.2020.69.3.474

[8] J. Pack and D. Kang, "RFDSys: Real-time Fire Detection System using AIoT-based K-NN and Motion Detection," *Journal of Korean Institute of Information Technology (JKIIT)*, Vol. 19, No. 10, pp.115-123, 2021. DOI: http://dx.doi.org/10.14801/jkiit.2021.19.10.115

[9] Y. Lee and J. Shim, "Deep Learning and Color Histogram based Fire and Smoke Detection Research," *International Journal of Advanced Smart Convergence(IJASC)*, Vol. 8, No. 2, pp.116-125, 2019. DOI: https://doi.org/10.7236/IJASC.2019.8.2.116

[10] J. Park, Y. Kim, "A Study on Deep Learning Performance Improvement Based on YOLOv5," in Proc. 78th the Korean Institute of Communication Sciences Conference, pp. 1592-1593, Jun. 22-24, 2022.

[11] C. Wang, H. M. Liao, I. Yeh, Y. Wu, P. Chen and J. Hsieh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," *in Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1-10, Jun.14-19, 2020.

[12] Y. H. Jeon, I. S. Kim and M. G. Lee, "Analysis of Results according to the YOLOv5 Model," *in Proc. Korean Society for Precision Engineering 2021 Autumn Conference*, pp.508, Nov.24-26, 2021.

[13] About mAP(mean Average Precision), https://ctkim.tistory.com/79

[14] T. Wang, S. Oh, H. Lee, D. Choi, J. Jang and M. Kim, "A Study on the Implementation of Real-Time Marine Deposited Waste Detection AI System and Performance Improvement Method by Data Screening and Class Segmentation," *The Journal of the Convergence on Culture Technology (JCCT)*, Vol. 8, No. 3, pp. 571-580, 2022. DOI: http://dx.doi.org/10.17703/JCCT.2022.8.3.571

[15] J. S. Kim and I. Y. Hong, "Analysis of Building Object Detection Based on the YOLO Neural Network Using UAV Images", *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vo1.39, No. 6, pp. 381-392, 2021. DOI: https://doi.org/10.7848/ksgpc.2021.39.6.381

[16] Roboflow, https://roboflow.com/