Implementation of a Fallen Person Detector

Junghak Kim and Geon-Woo Kim
Cyber Security Research Division
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
junghak@etri.re.kr

***Abstract*—This paper presents an implementation example of a detector that identifies fallen individuals as detection objects. The authors of this paper have been continually building datasets necessary for training and evaluating neural network models to be used as such detectors. This paper briefly explains the reasons why such datasets need to be constructed, and shows that, by utilizing theses datasets, it is possible to implement a detector capable of detecting human objects that represent the meaning of "fallen."**

Keywords—fallen person detection, object detection

I. Introduction

According to recent AI(Artificial Intelligence)-based video analysis technologies, determining whether a person has fallen on the street or somewhere else typically involves two main procedures. First, a procedure is carried out to detect objects identified as humans. Then, a subsequent procedure is carried out to analyze the features of each detected human object and determine whether it is classified as the category of "fallen". Alternatively, a procedure may be carried out to analyze and classify the consecutive action patterns or state changes of each human object, which is determined to be the same person across temporally consecutive video frames, by utilizing human action recognition technologies or similar techniques.

Object detection technologies, which detect a wide range of object classes, have reached a significant level of advancement over time. This can be attributed to the continuous advancement in the development of neural network architectures suitable for performing such tasks, as well as the methods for training those neural network models. However, it should not be overlooked that, behind this background, the construction of datasets for training those neural network models have played an important role considerably. A considerable number of datasets have been released in public and utilized for object detection tasks. But, nevertheless, these datasets also inevitably have clear limitations in covering all possible cases. Particularly, in the case of datasets related to human object detection, their scale and diversity may be considerably insufficient for implementing object detectors aimed at detecting human objects in atypical postures, such as those who are "fallen."

In relation to the above, this paper introduces the status of dataset construction required to detect human objects in a "fallen" state. And, this paper also introduces the results of implementing a human object detector utilizing the constructed datasets. In this paper, among well-known neural network models for object detection tasks, the training results for one of the latest YOLO (You Only Look Once) model is presented.

II. Implementation

A. Target Neural Network Models

It can be said that, among object detection technologies, YOLO is most popular technology. This is likely because the well-known framework released in public on GitHub allows anyone to easily and quickly implement an AI neural network model customized to their needs. The YOLO technology has been continuously advanced by numerous researchers through the development of neural network architectures and efficient training methods. As a result, it can be said that it has reached a considerable level in terms of detection accuracy and inference speed. Among the YOLO technologies, YOLO11, as a recently released technology, can be said to efficiently and appropriately extract the feature elements required for inference by effectively applying network layers responsible for contextual attention. In addition, another advantage of it is that it has slightly increased inference speed through improvements to the neural network architecture aimed at enhancing computational efficiency. The authors of this paper are conducting a research project on the development of an application framework that enables multiple neural network models to run simultaneously on mobile devices. In consideration of that, this paper presents the training results for YOLO11s, which is a light-weight model among YOLO11 models.

B. Datasets

The authors of this paper faced a situation where they had to make choices from the very beginning of dataset construction. Considering commonly used datasets, it was anticipated that building additional datasets to be integrated with existing ones according to their structures would require an enormous amount of time and manpower. Therefore, the authors determined that it was impossible to create a perfect datasets from scratch. So, as a preliminary step, labeling and bounding-box annotation were performed to detect only human objects which are classified as "fallen." Fig. 1 shows some examples of the annotated datasets.

(Note that these images are extracted from videos provided through AI-Hub.)

Fig. 1. Examples of the annotated datasets

TABLE I. DATASETS SCALE (THE NUMBER)

Source	Training			Evaluating		
	Annotated Images	*Annotated Instances*	*Background Images*	*Annotated Images*	*Annotated Instances*	*Background Images*
VFP290K	89,570	131,096	9,560	22,275	32,545	2,361
AI-Hub	94,134	94,254	135,233	23,303	23,330	33,499
Private	81,121	81,121	39,605	20,248	20,248	9,886
Total	264,825	306,471	184,398	65,826	76,123	45,746

The datasets constructed so far for detecting fallen person objects consists of images extracted from three different sources. The dataset construction began with the utilization of VFP209K datasets, which is organized to classify fallen person and non-fallen person into separate classes [3]. However, since there were quite a bit of annotation errors, only the images deemed necessary were selectively extracted, and the annotation work was performed again. Subsequently, the scale of the datasets were expanded by utilizing video datasets provided through AI-Hub, operated by the NIA(National Information Society Agency) of Korea, and this video datasets is still being utilized for the dataset construction for the fallen person detection. Additionally, the authors of this paper made private video datasets in which videos are recorded in apartment environments in Korea, and images extracted from these video datasets were also added to the datasets for the fallen person detection. Table I shows the scale of the datasets constructed so far.

C. Training

To train the YOLO11s model, basic functions such as loss function, optimization function, scheduling function, etc. were adopted according to the guidelines presented in [4]. However, there is one point to be noted in the process of training a model for fallen person detection. When processing images following the data augmentation process provided in [4], there are cases where the regions corresponding to the pre-annotated ground-truth bounding-boxes are partially cropped. While this may not cause a significant problem when training a model to detect the general object "person." But, in the case of detecting the object "fallen person," such partial cropping may lead to errors. This is because the partially cropped area of the annotated object may not contain sufficient information to determine that the object is "fallen." Therefore, to figure out this issue, the authors of this paper partially modified the data augmentation process of [4] so that the annotated bounding-box regions labeled as "fallen" are not cropped beyond a certain portion.

D. Results

Fig. 2 shows the training results. In Fig. 2, it shows that the curves representing the evaluation metrics converge in a stable manner. Consequently, as shown in Fig. 2, it can be said that it is possible to implement a "fallen person detector" capable of detecting only human objects identified as "fallen." In Fig.2, the value of mAP50-95, which is the mean average precision calculated at varying IoU(Intersection over Union), ranging from 0.5 to 0.95, goes up to about 0.99. However, this result only indicates that the target neural network model is well trained on the constructed datasets as intended. It does not mean that the neural network model trained on the constructed datasets can perform nearly perfectly in all cases beyond the scope of those datasets. Therefore, continuous dataset construction is required to ensure the target neural network model performs well across as many cases as possible.

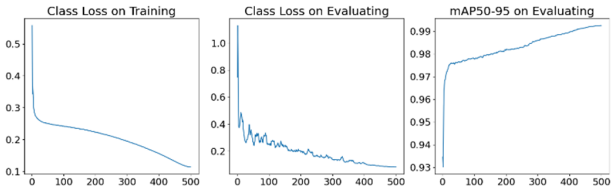

Fig. 2. Training results (Class loss and mAP50-95 curves)

To explain the reasons why these types of datasets need to be constructed, for the annotated images used in the evaluation process shown in Table I, the authors of this paper measured the success (recall) rate of detecting fallen person objects as human objects when utilizing the YOLO11s model officially release in [4]. Table II shows the above success rate according to some IoU thresholds. In table II, the IoU means the intersection over union between the ground-truth bounding-box and the predicted one by the official YOLO11s model. As shown in Table II, when a general object detection model released in public is used, there can be relatively high rate of failure to detect fallen people as human objects, which may often lead to situations where related video analysis systems fail to recognize fallen people.

TABLE II. RECALL RATE OF DETECTING FALLEN PERSON OBJECTS AS HUMAN OBJECTS WHEN UTILIZING THE OFFICIAL YOLO11S MODEL

Source	*Number of Annotated Instances*	*Recall Rate*				
		IoU ≥ 0.5	*IoU ≥ 0.6*	*IoU ≥ 0.7*	*IoU ≥ 0.8*	*IoU ≥ 0.9*
VFP290K	32,545	0.58	0.57	0.56	0.53	0.47
AI-Hub	23,330	0.67	0.63	0.59	0.54	0.41

III. CONCLUSION

In this paper, it's shown that it is possible to implement an object detector capable of detecting human objects that represent the meaning of "fallen." Of course, to figure out this type of task, the development of high-performance neural network models may be required. But, to enable such neural network models to perform their intended functions, it is sometimes inevitable that the construction of appropriate datasets must also be supported.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) [No. RS-2024-00394190 (Development of an Open Video Security Platform Technology with On-device Self-protection)].

REFERENCES

[1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," in Proceedings of the IEEE, Vol. 111, No. 3, pp. 257-276, March 2023.

[2] A. Salari, A. Djavadifar, X. R. Liu, and H. Najjaran, "Object Recognition Datasets and Challenges: A Review," arXiv:2507.22361 [cs.CV], 2025.

[3] J. An et al., "VFP290K: A Large Scale Benchmark Dataset for Vision-based Fallen Person Detection," NeurIPS 2021 (Thirty-Fifth Conference on Neural Information Processing Systems) Datasets and Benchmarks Track (Round 2), 2021.

[4] https://github.com/ultralytics

[5] https://github.com/DASH-Lab/VFP290k

[6] https://aihub.or.kr