ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

DEVELOPMENT OF MOBILE-DEVICE APP BASED ON YOLOV7 FOR SAFETY MONITORING

Jeong Tak Ryu*[1], BoKyeong Kim[2], JaeYeon Byun[3], and Kyung-Ae Cha[4]
[1]Dept. of Electronic Engineering, Daegu University , Gyeongsan, Republic of Korea
[2,3,4,5]Department of Artificial Intelligence, Daegu University Gyeongsan, Republic of Korea
*[1]jryu@daegu.ac.kr, [2]qhrud418@gmail.com, [3]9722jayon@gmail.com, and [4]chaka@daegu.ac.kr
*Corresponding author: jryu@daegu.ac.kr

ABSTRACT
This study introduces a fall detection system that uses the built-in cameras of mobile devices. Without the need for additional equipment or installations, users can access the proposed system through a mobile app. The system connects to a server via a network, storing data upon detecting a fall, issuing alerts to users, and monitoring incidents. This system is designed to be versatile, allowing it to be used in any location where individual users require such a system. This study commences with background research on intelligent monitoring systems and the related industrial trends. The overall system architecture is then discussed, including training data and models. The effectiveness of the system is verified through an implementation process and the introduction of performance metrics based on experimental applications of this system.

Keywords: Mobile app-based system, Real-time monitoring, YOLOv7

I. INTRODUCTION
Falls pose a high risk of causing severe injuries or even death, and can lead to negative outcomes such as fractures. Such incidents may result in loss of independence, disability, and psychological anxiety regarding the possibility of recurrence [1]. These incidents happen indiscriminately, regardless of time and place. While timely response to minimize harm in safety-related incidents is ideal, it is often missed. Immediate action is essential to prevent unexpected safety incidents. Consequently, the use of intelligent monitoring systems is on the rise in various situations, contributing to the growth of the related industries [2]. For example, IntelliVIX has grown beyond traditional CCTV to become a leading company in intelligent CCTV, achieving the highest domestic sales in this sector [2]. However, the installation costs for intelligent CCTV systems, including labor and materials, can run from 3 to 4 million won per unit for bulk orders, with the necessity to install multiple units driving up the overall expenses [3]. This presents a significant financial challenge for individuals, especially in homes which are the most common sites of safety incidents [1]. To overcome this, it is imperative to enhance the accessibility of intelligent CCTV systems with intuitive interfaces for home use.

In response to this need, our research has developed a safety monitoring system through a simple application that uses the built-in cameras of mobile devices, eliminating the need for additional equipment or installation processes. This system saves data on a server through network connectivity and provides users with real-time notifications and monitoring capabilities in the event of a fall. It is designed for universal application, suitable for use in homes as well as various external spaces as needed by individual users.

This paper is organized as follows: Section II explores the theoretical background of intelligent monitoring technologies and industry trends. Section III details the design of the overall system architecture and the development of necessary training data and models. Section IV addresses the actual implementation process of the system, and Section V presents experimental results and performance metrics to verify the system's effectiveness. Finally, Section VI concludes the study and discusses future research directions.

II. RELATED TECHNOLOGICAL TRENDS

A. Research Background Related to Industrial Trends
The technology behind intelligent image security is a next-generation technology that automatically analyzes images collected in real-time through a CCTV using both software and hardware functions designed to detect and

respond to anomalies. The main keywords in this industry include "4K CCTV," "artificial intelligence cameras," and "smart home security services." Intelligent CCTV technology based on deep learning has evolved from motion direction detection to the development of cloud sourcing [4]. Domestically, the primary trend of image security revolves around intelligent video analysis. Although CCTV initially gained prominence in securing public facilities and services, its applications have expanded to private sectors such as corporations and households. Furthermore, the demand for smart home services and CCTV systems is projected to continue expanding owing to the increasing proportion of single-person households and the surge in various types of crimes.

B. Object Detection via Artificial Intelligence

The You Only Look Once (YOLO) algorithm [5] is a deep learning-based approach for real-time object detection. YOLO divides the input image into a grid and simultaneously estimates bounding boxes and object probabilities for each grid cell. This processing method significantly improves computational efficiency by processing the entire image only once. Human recognition and fall detection require real-time decision-making and responses. Therefore, a learning model, i.e., YOLO, capable of accurate object recognition and fast processing was utilized to implement the system proposed in this study.

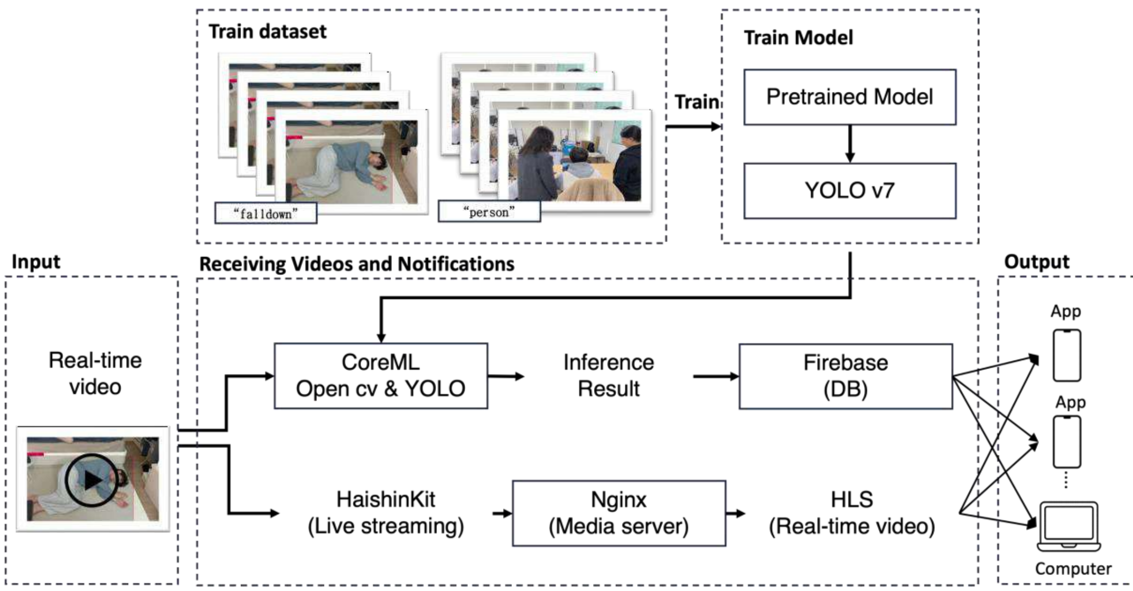

Fig. 1. System configuration

III. PROPOSED SYSTEM

A. System Structure

This study proposes a system for providing real-time event-detection alerts and video streaming, as shown in Fig. 1. The system integrates video data captured via smartphones with a trained YOLO model through the CoreML framework on iOS devices for inference. When specific events occur, the system sends users real-time notifications containing information such as location, time, detection circumstances, and detection images, enabling immediate responses.

Additionally, this system is designed with a structure that transmits video data to multiple mobile devices and computers through a media server. This architecture allows users to view the streaming video in real time across various devices.

The system proposed in this study can be flexibly applied in diverse environments, depending on the training data. This system can potentially be applied in various scenarios and settings, including home environments, elderly care and childcare, and industrial sites.

B. Dataset and Model Training

In this study, a learning model was constructed using the Roboflow image data labeling and sharing platform (Roboflow, Inc.) [6]. A total of 2,790 image data were processed on this platform and categorized into two classes: "person" and "falldown." Based on these two classes, two functionalities were implemented: "fall detection" and "person count detection."

Furthermore, 72%, 16%, and 11% of the dataset was divided into the training, validation, and test sets, respectively. To increase training speed and minimize the computing resource burden, all images were adjusted to a consistent resolution of 640×640. Data augmentation techniques, including image transformations, rotations, translations, flips, and resizing, were employed to increase the diversity of the training data. These transformations resulted in the construction of a total of 6,830 training images.

The model used in this study was trained based on YOLOv7 and subsequently converted to CoreML [7], a model format compatible with iOS, for usage. Coremltools was employed to convert the PyTorch model to CoreML such that the model can be used in the mlmodel format. For this study, Coremltools version 6.2.0 or below was used for the conversion.

The YOLOv7 model was used for training in this study. The training environment utilized the NVIDIA A100 graphics processing unit (GPU) provided by Colab. The hyperparameters included a total of 100 epochs, a batch size of 16, and an input image size of 640 × 640 pixels.

IV. IMPLEMENTATION

A. Design of the Proposed Monitoring System

Videos recorded using iOS devices offer two primary functions. The top within the "Receiving Videos and Notifications" box in Fig. 1 represents the intelligent CCTV mode that detects the situation presented in every frame from the input stream, i.e., whether a person falls down or too many people are within a particular area. When this mode is activated, in the event of a fall or if the number of individuals exceeds a predefined threshold, detection data is stored in Firebase, and notifications are sent to users.

The bottom within the "Receiving Videos and Notifications" box in Fig. 1 shows the second function, which provides the multi-platform streaming server to distribute real-time input videos from various types of devices. Using this streaming service, the app user can access on-time videos from the monitoring system.

This feature uses HaishinKit [8] to transmit videos at 30 frames/s in the real-time messaging protocol(RTMP) format to a Nginx [9] media server. On the receiving end, this video is processed in the hue, saturation, and lightness format with a .m3u8 file extension. When receiving real-time videos on iOS devices, the audio data must also be provided. Otherwise, the AV Player may not recognize this video as "live." As the video is received via Nginx, external access is possible for multiple users.

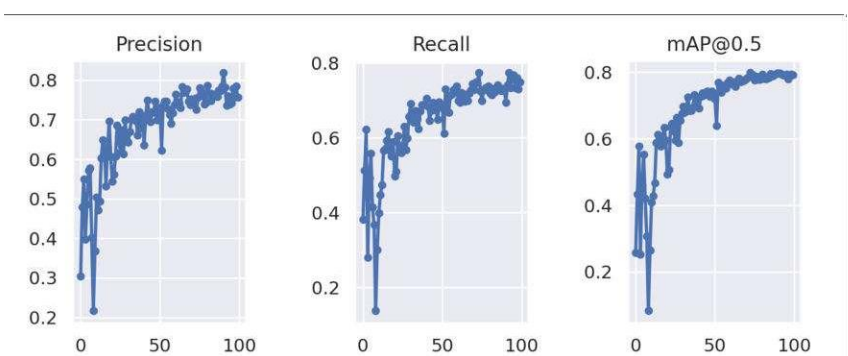

Fig. 2. Training result of the YOLOv7 model

Fig. 3. Example of a push message

Fig. 4. Alarm for incorrect number of workers at a job site

Fig. 5. Alarm for a fall incident at the job site

B. Detection Processing

In this study, individual frames of the input video were sliced using OpenCV, and each frame was inferred through CoreML. During this process, the obtained inference results contained detected class labels, and these values were then stored in a list. The primary points of interest in this study are the labels "person" and "falldown," which represent people and falls, respectively.

By aggregating the number of "person" and "falldown" labels stored in the list, the number of detected individuals and the occurrence of falls at specific points in time could be quantitatively assessed. Therefore, this system performed real-time object detection and quantitative analysis of outcomes, particularly related to person recognition and fall detection.

C. Prevention of Erroneous Detection

The preceding functions are aware that various variables can be generated during certain situations. For instance, a person may briefly exit the camera's view and re-enter or an object like a person may be erroneously detected as a person. To address such potential issues, counters were introduced for each function.

These counters include a fps counter that increments with each frame and function-specific counters that increment based on specific conditions, such as "when a fall is detected" or "when a certain number of people are detected." The fps counter resets after every 200 operations, whereas the function-specific counters become active after every 80 operations.

When the value of the function-specific counter exceeds a specified threshold, the corresponding information including time and location, is sent to Firebase. Subsequently, a warning notification is dispatched to the iOS device on which the application is installed. This enables flexible adaptation to dynamic elements in real-time environments.

V. EXPERIMENTAL RESULTS

A. Trained Model Performance

This section shows the performance of the model trained in this study and the experimental results.

The precision obtained was 0.7573, the recall was 0.7495, and the mean average precision (mAP)_0.5 score was 0.7912. The higher the mAP value, the better the performance.

As depicted in Fig. 2, the application developed in this study using the YOLO v7 model demonstrated excellent performance, even under challenging low-light conditions, by activating features such as fall detection and people counting. These results serve as evidence of the system's performance and practical applicability, as proposed in this research.

B. Application User Interface

The messaging service provided by Firebase can be used to send notifications with desired content to users. However, a drawback is that notifications can only be received when the application is actively running and not when the application is closed. To address this limitation, the use of the Apple push notifications (APNs) platform is necessary. Owing to Apple's security policies, if APNs are not used, notifications cannot be sent after the application has been closed. Therefore, when the application requests Firebase messaging functionality through APNs, users can receive notifications even if our proposed application is inactive. Fig. 3 presents an example of the notification text as a push message.

The implementation of the proposed application is visible in the displays on the user interface. Among its features, the zone verification function assists users in the real-time monitoring of specific areas. Fig. 4. The fall detection feature identifies abnormal movements, and Fig. 5. The people counting function automatically calculates the number of individuals in the video. Lastly, the notification feature delivers this information to users in real time.

Furthermore, the application developed herein securely manages individual user information by requiring users to undergo the Google login process when starting the application. The application also provides additional options and settings, allowing users to customize the system settings according to their preferences. Future updates aim to enhance the application further by incorporating various features and improving user convenience, ultimately achieving a more comprehensive application.

In addition, a low-resource mode was developed, by considering the performance differences between iOS devices. This mode efficiently separates real-time video streaming and model inference, thereby conserving computational resources. When the user is not actively monitoring the CCTV, only model inference is conducted. Real-time video streaming is initiated when the user accesses the detailed page.

VI. CONCLUSIONS

In this study, a portable monitoring system based on YOLOv7 real-time object detection is proposed. The objective of this system is to rapidly detect risks, provide warning notifications to users, and enable timely responses for accident detection. Future research will focus on performance improvements through various experiments and data enhancements, to advance the deep learning-based real-time monitoring system and improve the safety in various user environments.

Acknowledgement

This research was supported by Daegu University Research Grant, 2020

REFERENCES

[1] W. H. Abd, A. T. Sadiq, K. A. Hussein, "Human Fall Down Recognition using Coordinates Key Points Skeleton," *2022 3rd Information Technology To Enhance e-learning and Other Application (IT-ELA)*, Baghdad, Iraq, 27-28 December 2022, pp. 232-237, doi: 10.1109/IT-ELA57378.2022.10107951.G.U Kim "Special Report - Trend of Technology in Video Surveillance System," Telecommunications Technology Association, 2017.

[2] J. Park, K. Kim, "Comprehensive plan for national safety system reform," MOIS. Ministry of the Interior and Safety, February, 2023

[3] T.S. Jo, H.K. Kim, "Technical Committee Trend Report," Korea Conformity Laboratories,

[4] Wang, Chien-Yao & Bochkovskiy, Alexey & Liao, Hong-yuan. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the- art for real-time object detectors. 10.48550/arXiv.2207.02696.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[5] Roboflow, [Online]. https://roboflow.com/. [access: September 18, 2023]

[6] CoreML, [Online]. https://developer.apple.com/kr/machine-learning/core-ml/. [access: September 18, 2023]

[7] HaishinKit, [Online], https://github.com/shogo4405/HaishinKit.swift/. [access: September 18, 2023]

[8] Ngnix, [Online]. https://www.nginx.com/. [access: September 28, 2023]