Computers, Materials & Continua
DOI:10.32604/cmc.2022.023295
Article

Tech Science Press

Object Detection for Cargo Unloading System Based on Fuzzy C Means

Sunwoo Hwang[1], Jaemin Park[1], Jongun Won[2], Yongjang Kwon[3] and Youngmin Kim[1],*

[1]Department of Systems Engineering, Ajou University, Suwon, 16499, Korea
[2]New Transportation Innovative Research Center, Korea Railroad Research Institute, Uiwang, 16105, Korea
[3]Innovative Transportation and Logistics Research Center, Korea Railroad Research Institute, Uiwang, 16105, Korea
*Corresponding Author: Youngmin Kim. Email: pretty0m@ajou.ac.kr
Received: 02 September 2021; Accepted: 20 October 2021

Abstract: With the recent increase in the utilization of logistics and courier services, it is time for research on logistics systems fused with the fourth industry sector. Algorithm studies related to object recognition have been actively conducted in convergence with the emerging artificial intelligence field, but so far, algorithms suitable for automatic unloading devices that need to identify a number of unstructured cargoes require further development. In this study, the object recognition algorithm of the automatic loading device for cargo was selected as the subject of the study, and a cargo object recognition algorithm applicable to the automatic loading device is proposed to improve the amorphous cargo identification performance. The fuzzy convergence algorithm is an algorithm that applies Fuzzy C Means to existing algorithm forms that fuse YOLO(You Only Look Once) and Mask R-CNN(Regions with Convolutional Neuron Networks). Experiments conducted using the fuzzy convergence algorithm showed an average of 33 FPS(Frames Per Second) and a recognition rate of 95%. In addition, there were significant improvements in the range of actual box recognition. The results of this study can contribute to improving the performance of identifying amorphous cargoes in automatic loading devices.

Keywords: Deep learning algorithm; YOLOv2; Mask R-CNN; Fuzzy C Means; unloading system

1 Introduction

Artificial intelligence and deep learning technologies have recently developed faster than in the past few centuries, and with the development of these technologies, various deep learning algorithms are being used in the field of object and pattern recognition. Object and pattern recognition fields can be utilized in a number of fields, such as military use, autonomous driving, and consequently eliminate human physical fatigue and cognitive errors that can occur in those fields. In the logistics sector, it is difficult to apply the object and pattern recognition field to handle unstructured cargoes of different shapes and sizes. With the recent the growth of the online market, the logistics center is investing in developing innovative technologies to handle the increased volume. The automatic unloading system for improving cargo handling efficiency must requires a deep running algorithm technology that can recognize objects. The development of algorithms related to object recognition is actively underway. However, algorithms that are suitable for automatic unloading systems that require the identification of large numbers of unstructured cargo need to be developed. This paper aims to propose a deep learning algorithm for the recognition of cargo objects in automatic unloading devices. To solve the problem, we would like to analyze the deep learning algorithm and prior research and develop an applicable improvement plan. Fig. 1 shows a schematic view of the courier cargo automatic loading device considered in this study.

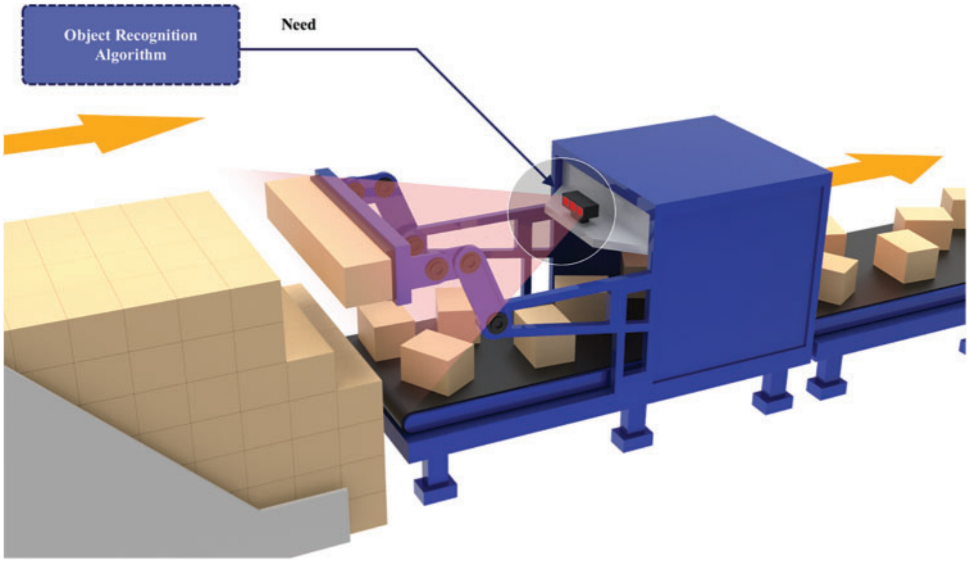

Figure 1: Automatic cargo unloading system

1.1 Related Literature Review

In order to implement efficient logistics processes, it is essential to develop robot technology that combines various fields, and underlying research using various approaches has been conducted. Based on the relevant technology and case study analysis results, a new cargo object recognition algorithm is proposed. Shin et al. conducted a study on the utilization trends of logistics technologies based on the 4th Industrial Revolution to examine the current status and implications of the logistics industry of 4th Industrial Revolution technologies such as robots, blockchain, Internet of Things, and big data. As a result, we concluded that for artificial intelligence, combined with robots, the existing workforce is well replaced and blockchain-based smart contracts will achieve the efficiency of the logistics process. In addition, it was argued that this requires a supporting and educational infrastructure to train and demand for personnel with artificial intelligence, big data analysis capabilities [1]. Kwak et al. conducted tracking and spatial operation of facilities, components, finished products, etc. in simulation-based manufacturing sites on methods for smart SCM(Supply Chain Management), including design and implementation of logistics SCM system in a research on logistics object tracking service for smart SCM. The study concluded that by combining detailed technology with logistics units, unnecessary waste of resources and object tracking, technology improvement can be contributed to enhancing corporate productivity by various manufacturing units such as quality control and product production, packaging and delivery [2]. Yu et al. understood trends in robot work intelligence technology in the intelligent logistics/agriculture field and conducted research on robot work intelligence for automation of logistics, robot work intelligence for automation of agriculture, and concluded that standardization for interfaces between heterogeneous technologies and robots should be carried out [3]. In the 4th Industrial Revolution, Choi et al. identified trends in the driving/manipulation technology of logistics robots, technology trends of delivery robots, and related element technologies. Cameras were sensitive to changes in illumination due to time and weather, and concluded that research on these areas was needed in many ways, from accurately recognizing/estimating the shape of objects to capturing unknown shapes of objects by robots [4]. In a study of automatic picking/classification system using image analysis, Park et al. produced a simple and repetitive type of equipment that performs picking/classification operations on industrial sites. This study derived the meaning as a prototype because it implements control using the Communications Department, although it is a reduced-scale picking/classification equipment [5].

Won et al. analyzed deep learning algorithms in deep learning-based cargo recognition algorithm studies for courier cargo automatic unloading equipment, and proposed deep learning algorithm models that added a masking network that increased the accuracy of bounding boxes to YOLOv2 model base [6]. In a study on the factors affecting the intention to purchase logistics robots in the logistics center through the technology acceptance model, Hwang et al. prepared 11 hypotheses on technology acceptance variables and conducted an analysis on the results of the survey. The purpose of the study is to understand the factors affecting the purchase intention of logistics robots in order to be applied to domestic logistics centers, and to spread logistics robots, it is necessary to focus on the usability of logistics robot technology and establish strategies to increase its usefulness [7]. This paper is based on the development of ICT (Information and Communication Technologies) technology related to the 4th Industrial Revolution, and applied to the automatic unloading system by incorporating object recognition algorithms in the field of artificial intelligence deep learning into logistics 4.0. Fig. 2 shows a conceptual diagram for this.

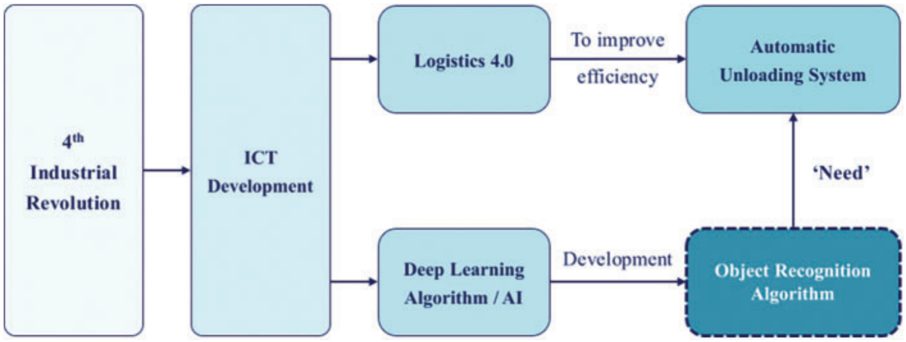
4th Industrial Revolution
ICT Development
Logistics 4.0
To improve efficiency
Automatic Unloading System
'Need'
Deep Learning Algorithm / AI
Development
Object Recognition Algorithm

Figure 2: Development of automation equipment and necessity of algorithm

Artificial intelligence technology to simulate human intelligence has been gradually developed. This artificial intelligence technology attracted a lot of public attention when Deep Blue beat the world chess champion G. Kasparov in 1997. While this deep blue was successful in many aspects of popularity, Deep Blue's intelligence was technically difficult to appreciate because it relied on the knowledge of digitized chess masters and computing power to calculate the vast number of cases. By comparison, recent AI technologies, represented by IBM's Watson, Apple's Siri, Google Now, and others, have advanced toward human-level AI in that they automatically accumulate knowledge based on big data instead of relying on expertise in the field. There are many perspectives and approaches to understanding artificial intelligence, but let's take a look at data-based artificial intelligence, which has led the rapid improvement in pattern recognition performance recently. Despite theoretical advances in pattern recognition based on small data in the 1990s, its performance fell far short of that of human intelligence. Deep learning is an important artificial intelligence technology that leads to performance improvements in various pattern recognition fields, including voice recognition and image recognition, in this context. There is a lot of research going on how to create better expression techniques and models to learn to present it in a form that computers can understand when there is any data and as a result of this effort. The concept of deep learning has actually been proposed and discussed in the 1980s or earlier. However, only after Professor Hinton's 2006 paper published in Science did many people begin to systematically study deep learning [8]. There have been several attempts to increase the computational ability of deep learning, one of which was to create more neural network of deep learning. On the contrary, there was also a discussion that creating a lot of deep learning neural network networks has limitations in enhancing computational capabilities. It was the emergence of backpropagation algorithms published by D. Rumelhart, G. Hinton, and R. Williams in 1986 that people began to be enthusiastic about neural networks again.

In fact, backpropagation algorithms had been around before, but they began to draw attention from their papers in 1986, and the neural network again attracted people's attention with optimistic prospects. This backpropagation algorithm has made not only single-layer neural networks but also multi-layered neural networks with one or two hidden layers learnable. However, when SVMs(Support Vector Machines) were introduced by V. Vapnik and C. Cortes in 1995, and performed better than neural networks, people abandoned the neural network and flocked to SVMs. Over the next decade or so, the neural network has been subjected to researchers' indifference and neglect, but it has started to draw attention again based on Professor Hinton of the University of Toronto's 2006 Science paper. Currently, the paradigm of pattern recognition has changed in the artificial intelligence field, and research on voice recognition and image recognition is being conducted. In addition, it is recognized as a technology that can mature the level of artificial intelligence by making achievements in areas such as language comprehension. Currently, the most commonly used deep learning models for various pattern recognition competitions and use services are CNNs(Convolutional Neural Networks) and RNNs(Recurrent Neural Networks). CNNs' first computational model is the Neocognitron published by Fukushima in the 1980s [9]. Later in 1989, Y. LeCun combined the backpropagation algorithm with Neocognitron to create CNNs. Recently, the trend of image recognition is to maximize performance by expanding the size of CNNs and designing them to have various structures. For RNNs, LSTM(Long Short-Term Memory), a type of RNNs, has recently been successfully applied to cursive recognition or speech recognition as a neural network for time series data analysis [10]. Despite the successful operation of the algorithm, neural network learning took nearly three days and this was considered unrealistic to be generally applicable to other fields. Nevertheless, there are three main reasons why deep learning has been revived. The first is that the drawbacks of existing artificial neural network models, which have previously been mentioned in the history of deep learning, have been overcome. For a second reason, there is another factor in hardware development. In particular, powerful GPUs significantly reduced the time spent on complex matrix operations in deep learning. Finally, the third most important reason is Big Data. The massive influx of data, and efforts to collect them, can all be aggregated, analyzed and used for learning, especially large amounts of data and tag information produced by SNS users. Training vectors used for learning in artificial neural networks should be labelled data (for supervised learning), and it is impossible to label all large training sets. For this reason, supervised learning is performed only on some of the data used for initial learning and unsupervised learning is performed on the rest of the training sets, and the results learned combine the results of the existing learning and the meta-tag information analyzed earlier to complete the recognizer. Since the resurgence of deep learning, we have seen the highest levels of performance in various fields, especially in the field of ASR(Automatic Speech Recognition) and computer vision, which are typically TIMIT(Texas Instruments and MIT-generated voice databases) and MNIST(Modified National Institute of Standards and Technology database). Recently, deep learning algorithms based on Convolution Neural Networks have shown excellent performance, especially in areas such as computer vision and voice recognition. MNIST database data is typically used as evaluation data for image classification. MNIST consists of handwritten numbers, including 60,000 learning examples and 10,000 test examples. Similar to TIMIT, low-capacity MNIST data enables multiple test configurations. This has led to the importance of deep learning in the areas of image recognition and object recognition, which are major areas of computer vision. At that time, knowing that deep learning worked fairly well for large-scale speech recognition, they used a deep convolutional neural network structure designed 20 years ago on a large scale to fit large-scale tasks. From 2013 to 2014, the error rate of ImageNet task results using deep learning quickly decreased, coinciding with the trend in the large speech recognition field. As with the expansion of the field of automatic speech recognition into the field of automatic speech translation and understanding, the field of image classification has expanded to a more challenging field called automatic image captioning.

1.2 Methodology

Computer vision and video processing cover a series of processes that process and analyze image data, such as photography and video, to extract information embedded in the data. Typically, it classifies classes of objects in photographs or videos, or detects the location of objects, as well as includes areas that address integrated cognitive-based interaction problems coupled with other branches of artificial intelligence, speech recognition, natural language processing, etc. The image consists of a distribution of dots in a two-dimensional space called pixels. Traditional methods digitized the correlation between pixels in two-dimensional space to screen highly correlated regions and call them features. Recently, it has been utilized for a variety of purposes, including recognizing objects based on distance information and applying them to moving robots [11]. Object detection is a more challenging problem because it estimates the class of objects in the image as well as the positional information of the objects. In particular, there are examples of using multiple background models to perform object detection with techniques that are drawing attention in the field of embedded machine vision related to security and surveillance systems [12]. The most widely used object detection methods in relation to machine vision algorithms are local feature matching-based methods. A typical regional feature matching process selects easy-to-identify features such as Harris corner, and extracts feature vectors from local patches around the selected feature points. Local scale-invariant feature point detection methods such as SURF are widely used to extract this feature vector [13,14]. In this regard, Pedro F. Felzenszwalb et al. conducted the Object Detection with Discipline Trained Part Based Models study [15]. The problem with this method is how to build a valid object model and the slow computation of complex feature vector and matching relationship analysis processes, which makes it difficult to apply in real-world environments. Not only is it difficult for humans to design appropriate features for images in all domains, but in terms of overall performance and efficiency, alternative measures have begun to be required for machines to derive appropriate features from data themselves. The recent rapid growth of deep learning technologies has marked a new turning point, proving to replace most of the problems involved in feature detection. Demand has exploded in a variety of areas, ranging from fine-precision tasks such as detection of defects in automated factory systems to comprehensive interaction tasks based on image information such as recognition of emotions between humans and robots.

Traditional object detection studies utilized methods such as SIFT(Scale-Invariant Feature Transform), SURF(Speed-Up Robust Feature), and HOG(Histogram of Oriented Gradient) based on low-level features. These approaches have faced limitations in performance improvement and various attempts have occurred to utilize CNN to detect objects since 2012. Based on the Selective Search algorithm, R-CNN proposed by Girshick's research team proposes a region pro-position in the input image. For each region, we classify the class of object patterns in the region through the CNN and through the SVM classifier [16]. This CNN is based on AlexNet and imports a model of a preferentially learned state for object classification. Mask R-CNN can be viewed as an algorithm that synthesizes the RPN(Region Proposition Network) and Fast R-CNN, and a new mask branch has been added to Fast R-CNN's classification, localization (binding box regression) branch. It is also an algorithm in which RoI(Region Of Interest) align replaced RoI pooling for FPN(Feature Pyramid Network) before RPN, and masking of image segmentation. The Mask R-CNN first resizes the image and is then input-sized to enter the backbone network. Create a featuremap on each layer via ResNet-101, and an additional feature map from a feature map previously created via FPN. Each RPN is applied to the final generated feature map to generate an out result. Among the anchor boxes created by Non-max-suppression, delete all except the anchor boxes with the highest score, and size the anchor boxes with different sizes through RoI alignment. Finally, pass the anchor box value to the mask branch to finish the final process. As a related study, Qiqiang Chen performed road damage detection and classification studies using dense Mask R-CNNs [17]. For road damage detection and classification, DenseNet produced better results than conventional algorithms in its first attempt to apply the Mask R-CNN framework. Ruohan Meng conducted a new steganography study based on instance segmentation [18]. A method based on instance segmentation such as Mask R-CNN has been proposed.

Currently, most deep learning-based object recognition and detection have addressed this by constructing the process of estimating the object's location area and classifying the object's class using a separate CNN-based deep learning model. Object detection models in the R-CNN family first estimate the object candidate area in the image and then find class class classification and object boundary boxes based on it. In this process, due to the large number of estimated candidate areas and their resulting overhead, there is a limitation to their performance as an application for utilization in the field, such as real security systems or robot remote control, in terms of detection rates. Therefore, various studies have been attempted to improve detection speed while maintaining object recognition rates. Among them, the YOLO network, which has recently received the most attention, is designed to perform both boundary box detection and class classification at the final output stage. In contrast to the three modules that make up the network in the fastest and most accurate Fastest R-CNN family models, which are responsible for feature detection, boundary box generation, and class classification, YOLO is simple and fast because all steps take place within a single network. The final output terminal of the YOLO network is a feature tensor that represents all potential candidates for object classes and boundary boxes in the image. This tensor divides the input image into a grid of a certain size and expresses the post probability that the boundary box generated in each grid will be the boundary box for the target object and what the class of the object is. Measurements of accuracy show somewhat lower results compared to Fast R-CNN, but offset this because it is overwhelmingly fast in terms of speed. YOLO, a type of deep learning algorithm for object recognition, divides the image to be predicted into grid cells and predicts one object for each cell. A predetermined number of Boundary boxes determine the location and size of objects, and since only one object can be predicted for each cell, the effectiveness of multiple objects overlapping can be compromised. Each boundary box consists of the location (x, y), size (w, h), and box confidence scores, with a total of five factors. The box confidence score reflects the probability that the box contains objects and how accurate the boundary box is. Conditional class probabilities are probabilities of which particular class the detected object belongs to. YOLO predicts multiple bounding boxes for each grid cell, but must select one box that best contains detected objects to calculate loss for true positives. To do this, calculate ground truth and IOU and select one with the highest IOU. This results in better predictions for size or aspect and aspect ratios. As a related study, Chintakindi Balaram Murthy conducted an efficient pedestrian detection algorithm study using the YOLOv2 model [19]. In this work, K-means clustering techniques were applied to the Pascal Voc-2007 + 2012 pedestrian dataset, and the fuzzy convergence algorithm showed more effective results in detecting pedestrians than conventional algorithms. Javaria Amin conducted a Convolution Bi-LSTM-based human pedestrian recognition study using video sequences, resulting in 90% accurate prediction on the CASIA-A, CASIA-B, and CASIA-C datasets with the fuzzy convergence algorithm [20]. In addition, Javaria Amin conducted a 3D semantic deep learning network study for leukemia detection using YOLOV2, utilized ALL-IDB1, ALL-IDB2 and LISC datasets to verify the precision, accuracy, and sensitivity of the fuzzy convergence algorithm, and produced more effective results [21]. Tab. 1 shows characteristic comparisons for YOLOv2 and Mask R-CNN. Qingyang Zhou recognized the problem of the existing object recognition algorithm that small objects cannot be accurately detected when there is an obstacle around them in the safety helmet detection method, and conducted a helmet wearing detection algorithm study based on AT-YOLO deep mode. Experiments have shown that the mAP of the proposed method can reach 96.5%, and the detection rate can reach 27 fps, Compared to other existing methods, detection accuracy and speed are superior [22]. Asma Baccouche proposed an end-to-end system based on a YOLO (Only-Look Once) model to simultaneously localize and classify suspicious breast lesions in mammography, and evaluated the model on two publicly available datasets with 235 mammograms in the INbrest database [23]. As a result of the study, the detection accuracy of 95.7%, 98.1%, and 98%, and 74.4%, 71.8%, and 73.2%, respectively, for mass lesions and calcified lesions in CBIS-DDSM, INbreast and personal data sets.

Table 1: Fuzzy convergence algorithm recognition process

Algorithm	Characteristic	Feature	Limit
Mask R-CNN	Rapidity	Real-time object detection available	Pixel unit detection difficulty
YOLOv2	Accuracy	Information acquisition possible without data labels	Rapid object detection difficulty

Deep learning is a powerful machine learning technique that can be used to training object detectors. Algorithms of YOLO and R-CNN series are commonly used, and YOLO's developmental type, YOLOv2, is capable of rapid detection with real-time object detectors [24]. The advanced version of the R-CNN family, Mask R-CNN, has the advantage of separating objects and obtaining data even if the data label is not displayed. The case study also identified that FCM(Fuzzy C Means) based on the Fuzzy function could be a solution if the object recognition process is uneven or the pattern boundaries are not clear [25]. FCM application has identified the ability to increase three-dimensional object recognition performance. Kim et al. conducted a pattern segmentation study using FCM, recognizing the problem that pattern segmentation is difficult if images are uneven or pattern boundaries are not clear [26]. As a result of the study, we propose a novel algorithm for detecting image patterns from early face pattern images to accurately measure three-dimensional face information using spatial coding patterns. The proposed pattern segmentation method as a result of the study showed improved segmentation efficiency over conventional methods. The objective function of clustering is expressed as shown in Eqs. (1) and (2).

$$\mathrm{J}(u_{ik},\ v_i) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} (d_{ik})^2 \tag{1}$$

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^{l} (x_{kj} - v_{ij})^2 \right]^{1/2} \tag{2}$$

where u_{jk} is the degree of belonging to x$_k$'s kth data in the Ith cluster, with a value between 0 and 1. When applied to image segmentation, corresponds to the pixel value of each pixel. And vi is the Ith cluster centroid vector, corresponding to the pixel pixel value of each cluster. j (j = 1, . . . , l) is a variable in the characteristic space, and m is the weight of the index indicating the effect of the degree of fuzzification of the belonging function. The results of the purpose function, Eqs. (1) and (2), can be expressed in Eqs. (3) and (4).

$$v_{ij} = \frac{\sum_{k=1}^{n} (u_{ik})^m x_{kj}}{\sum_{k=1}^{n} (u_{ik})^m} \tag{3}$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m}-1}} \tag{4}$$

Utilizing the presented equations, we perform the process of calculating the center of the fuzzy cluster and calculating a new belonging function. The equation for deriving the belonging function can be expressed by Eq. (5). Using Eq. (6) to derive the belonging function and Eq. (6) to derive the amount of change in the threshold, the amount of change in the threshold is determined. If the amount of change in the threshold is not appropriate, the algorithm is repeated until an appropriate threshold is derived by applying r = r + 1.

$$u_{ik}^{(r+1)} = \frac{1}{\sum_{j=1}^{c} \frac{(d_{ik}^{r})}{(d_{jk}^{r})}^{\frac{2}{m}-1}} \tag{5}$$

$$\triangle = \| U^{(r+1)} - U^{(r)} \| = max_{i,\,k} \left| u_{ik}^{(r+1)} - u_{ik}^{(r)} \right| \tag{6}$$

In this study, we propose an algorithm to improve the performance of object detection by selecting a delivery top-to-bottom robot in the logistics field as the target of the study. We perform technical analysis and case study to review the methodology needed to solve the problem and propose an algorithm with improved performance by improving the existing algorithm. To analyze the performance of the fuzzy convergence algorithm, verification experiments were conducted by implementing an experimental environment similar to the actual environment to derive results. Fig. 3 illustrates the performance of this study.

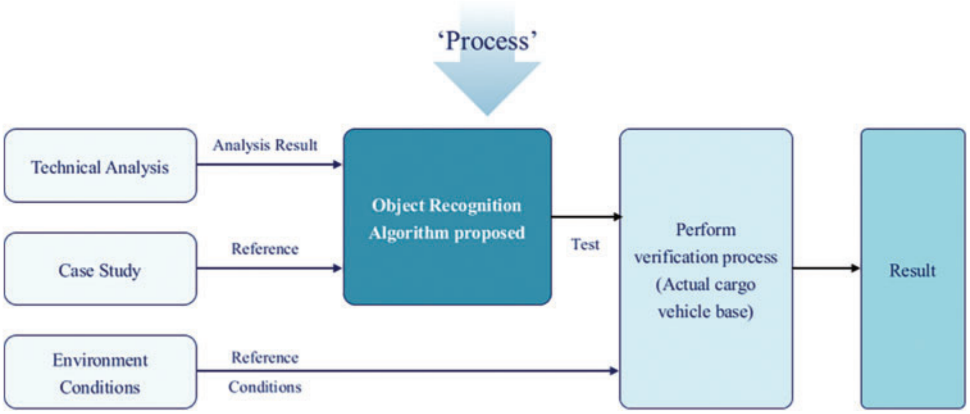

Figure 3: Fuzzy convergence algorithm development process

2 Fuzzy Convergence Algorithm

The YOLOv2 algorithm, which is used as an object recognition method in the existing deep learning field, has an advantage in speed and can be used for real-time object detection, but it is difficult to detect pixel-wise images. Furthermore, the Mask R-CNN algorithm has the advantage of accuracy, and although it is possible to analyze images to gather information without data labels, it is difficult to quickly detect objects. Fig. 4 outlines the problems with this, and shows a schematic process for it. Fuzzy C Means is combined with the speed of YOLOv2 and accuracy of Mask R-CNN to increase certainty about object recognition. The problem with these existing algorithms can cause errors in the logistics sector that do not properly detect couriers or calculate exact ranges even if they detect couriers with unstructured shapes. In addition, the processing speed of the calculation process is significantly slow, which may not be utilized in the real world. As such, the algorithms presented so far have problems in various aspects, and it is still difficult to use them for object recognition of unstructured parcels in the logistics industry. In this paper, we intend to improve the courier cargo object recognition performance of logistics top-to-bottom robots by combining the existing deep learning algorithms YOLOv2 and Mask R-CNN algorithms with the Fuzzy C Means methodology.

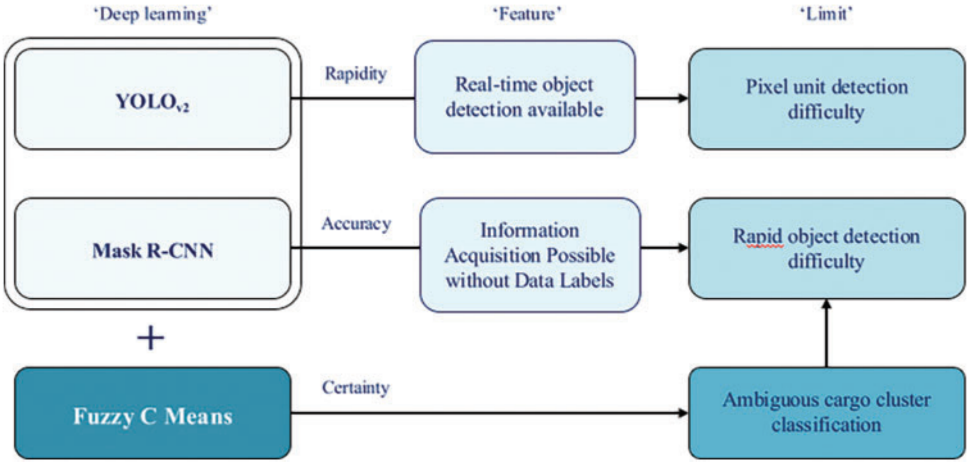
'Deep learning'
'Feature'
'Limit'
YOLOv2
Rapidity
Real-time object detection available
Pixel unit detection difficulty
Mask R-CNN
Accuracy
Information Acquisition Possible without Data Labels
Rapid object detection difficulty
+
Fuzzy C Means
Certainty
Ambiguous cargo cluster classification

Figure 4: Characteristics of YOLO and Mask R-CNN

The deep learning technology for cargo recognition proposed in this study uses an improved YOLOv2 technology. Because cargoes are usually loaded with regular cargoes (boxes) and unstructured cargoes, RGB images with no space on the bottom of the object (margin on the image), the typical R-CNN family deep learning technology has problems recognizing the location of the cargoes when large-sized errors occur. Therefore, in this study, we propose a model that extracts objects based on the YOLOv2 model characterized by fast real-time object detection and adds a masking network that increases the accuracy of bounding boxes to the object. Fig. 5 shows the neural network of the fuzzy convergence object detection algorithm.

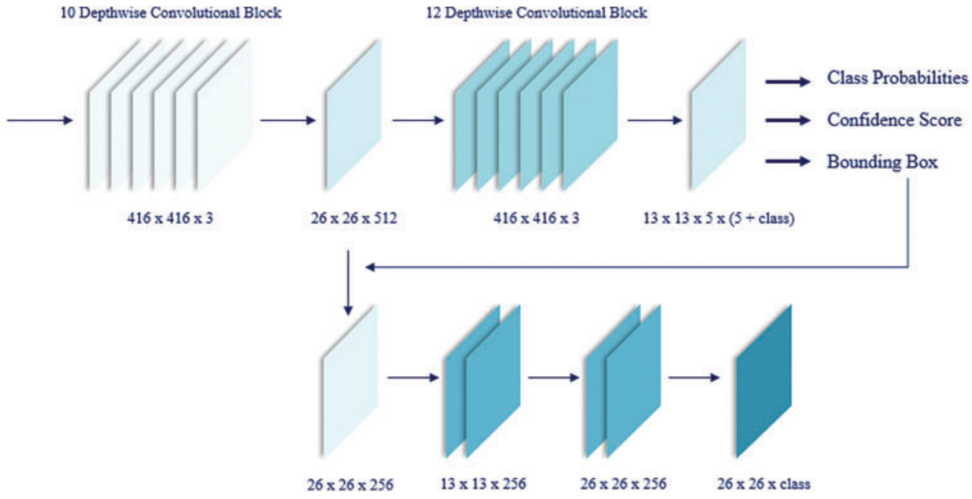

Figure 5: Diagram of the fuzzy convergence algorithm

3 Validation and Verification Through Experiment

3.1 Experimental Configuration

To verify the fuzzy convergence algorithm, we implement an experimental environment similar to the actual courier top-and-bottom system. Fig. 6 shows a schematic diagram for algorithm verification. The courier cargo automatic loading system consists largely of a courier cargo transport unit, multiple loading manifold, multiple loading equipment transport unit, incoming module, variable conveyor, and suction gripper. Adsorption grippers can be adsorbed up to 15 cm at a time when applying linear springs for pick-up of courier cargo and adsorbing cargoes with different depths. The frame of the adsorption gripper and the multi-loading manifold is designed to be 2 m high and 2 m wide, taking into account the height and width of the actual courier vehicle, and attaches a stereo camera to the top. In order to handle cargo at the front using one stereo camera, the camera position is equipped with a 2 m high manifold frame and the angle is adjusted between approximately 15 and 20 degrees. If the adsorption gripper consists of 20 adsorption modules of 10 cm∗10 cm, each adsorption module can be individually controlled. Using stereo cameras and deep learning algorithms, the first single-carrier cargo on the front is recognized and the work space and schedule are determined according to the size of the adsorption gripper. According to the determined work schedule, the courier cargoes of the topmost work space are adsorbed and unloaded on the variable conveyor, and when the first stage of work is completed, the multi-loading equipment transfer unit moves the manifold to about 2 m away from the next stage. During movement, the deep learning algorithm recognizes the location and distance of the courier cargo and controls the manifold by determining the next work schedule.

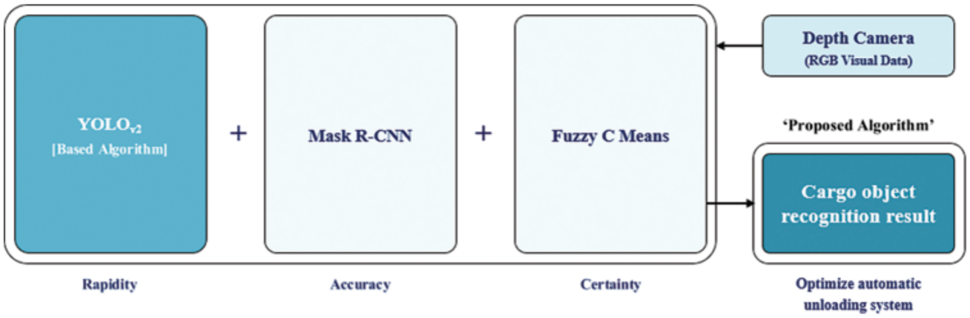

Figure 6: Configuring the fuzzy convergence algorithm verification environment

The schematic system configuration of these courier cargo automatic loading devices and the process for achieving their objectives were identified and reflected in the algorithm verification experimental environment. The distance between the stereo camera that recognizes the courier cargo image and the target courier cargo was kept at 1.5 m, and the stereo camera was positioned to recognize the courier cargo at a height of 2 m. In addition, to increase the efficiency of recognition, LED (Light-Emitting Diode) lights were installed on stereo cameras to perform verification of courier cargo recognition algorithms. Fig. 7 shows the composition of the actual verification experimental environment.

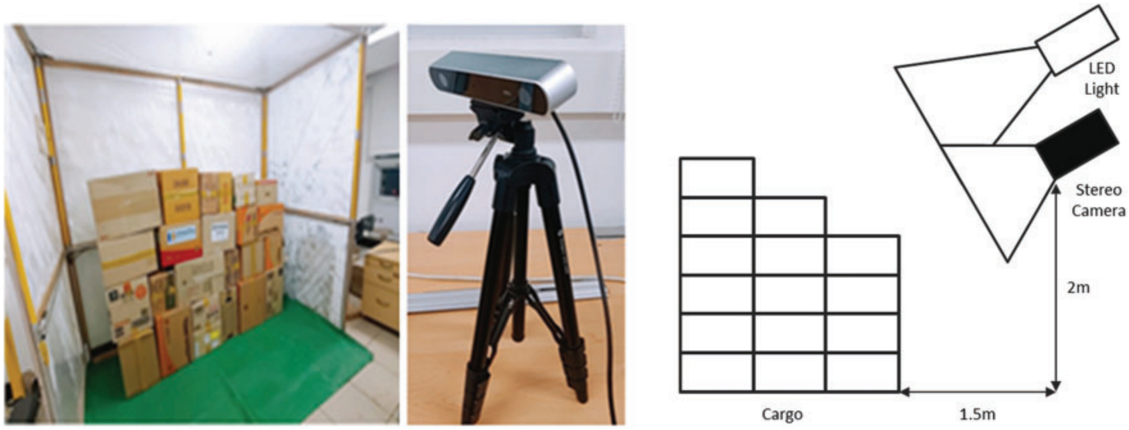

Figure 7: Verification environment configuration

3.2 *Experimental Environment*

The depth histogram is calculated by applying a histogram to the depth value of the stereo camera. The interval of the histogram is implemented at intervals of 20 mm. The distinct frequency for interval 20 mm shall be at least a certain value to be representative of the interval, and the representative values shall be at least 140 mm deviated and divided into up to three layers. Here, a constant value is given in 2∗ (width∗length) if it can vary depending on the resolution. This value is the frequency of the depth values of non-cargo background values, and those below these values are considered backgrounds, not cargoes. Among the valid representative values, the representative values present at the most distance represent the average distance value of the boxes stacked nearest. Use deep learning to recognize the boxes and extract the center points of individual boxes. Clustering between boxes at height (y-value) based on extracted center points, center points closer than 200 mm based on y-value of the top center point are designated as the first block, and center points closer than 200 mm based on the top center point are designated as the second block. Tab. 2 shows the set values of algorithm performance, and Fig. 8 shows the experimental process step by step.

Table 2: Composition factor of fuzzy convergence algorithm

Hyper parameter	Setting value	Setting method
Initial learning rate	0.001	Variation with optimization functions
Learning rate variation ratio	10%	Fix
Batch size	8	Fix
Total epoch	100	Fix

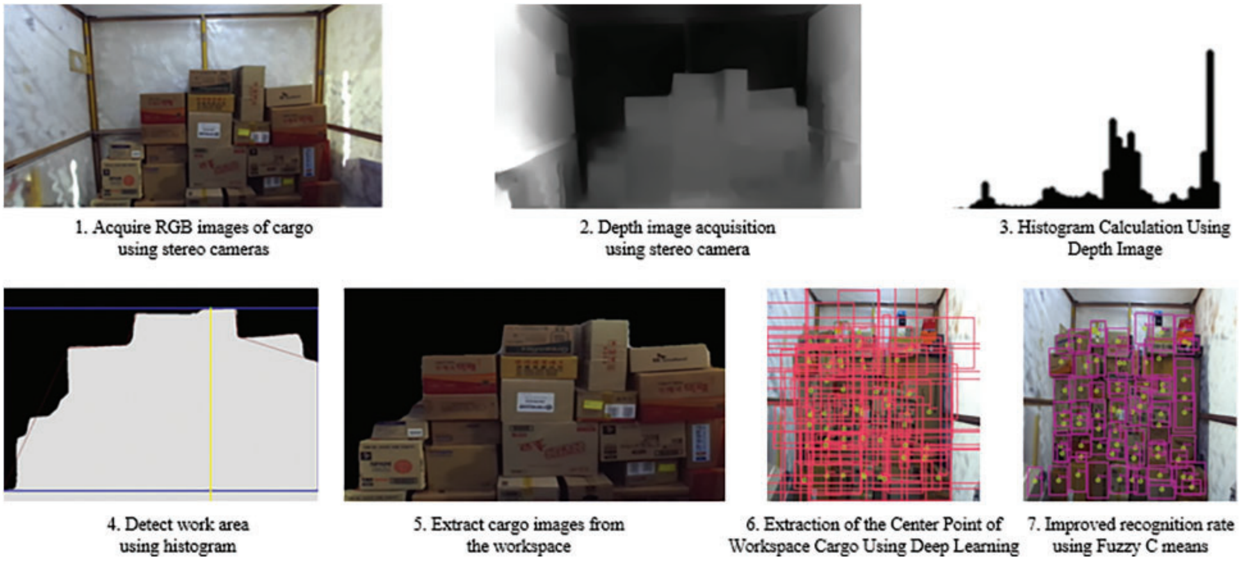
1. Acquire RGB images of cargo using stereo cameras
2. Depth image acquisition using stereo camera
3. Histogram Calculation Using Depth Image
4. Detect work area using histogram
5. Extract cargo images from the workspace
6. Extraction of the Center Point of Workspace Cargo Using Deep Learning
7. Improved recognition rate using Fuzzy C means

Figure 8: Fuzzy convergence algorithm recognition process

4 Result Analysis of Fuzzy Convergence Algorithm

To evaluate the performance of the fuzzy convergence algorithm, a space similar to the actual application environment was constructed, and the experimental space was based on the loading space of the container cargo vehicle. The test data were required to include both single box images and loaded box images. We also evaluated the performance over the number of batch sizes that could affect the performance of the proposed model. Testing Data and verification data were learned using 4,000 images, respectively. As a result of the experiment, when batch size was 8, the recall rate was 8.32 and the courier service recognition was the best. In RGB image data, we show images of the overall detection and recognition sequences described above, and show object ejection and center point results using deep learning. Furthermore, the proposed aneurism demonstrates satisfactory performance in the detection and recognition of courier cargoes in real-world experiments. The performance of the fuzzy convergence algorithm was verified by building a laboratory space similar to the actual work space. Using a number of cargo boxes, 4,000 sheets were learned from the algorithm each to conduct the verification. The fuzzy convergence algorithm and individual algorithms were validated together under the same environmental conditions for performance comparison. The experimental parameters and results can be found in Tabs. 2 and 3. Tab. 3 shows the results of the algorithm verification performance. The FPS average of the existing Mask R-CNN algorithm is 13, the number of objects detected is 48, the FPS average of the YOLOv2 algorithm is 45, and the number of objects detected is 44. The fuzzy convergence algorithm has an FPS average of 33, and the number of objects detected is 47. When comparing each algorithm, the fuzzy C mean convergence algorithm has a lower FPS mean than YOLOv2 and a lower recognition rate than Mask R-CNN. However, the fuzzy C mean convergence algorithm compensated for the shortcomings of the two existing algorithms and showed high cargo recognition bounding box density of the existing algorithms. Fig. 9 shows the result s for this.

Table 3: Fuzzy convergence algorithm recognition process

Model	FPS Average	Actual number of boxes	Number of recognized boxes	Recognition rate
Mask R-CNN	13	50	48	96%
YOLOv2	45	50	44	88%
Fuzzy convergence algorithm	33	50	47	95%

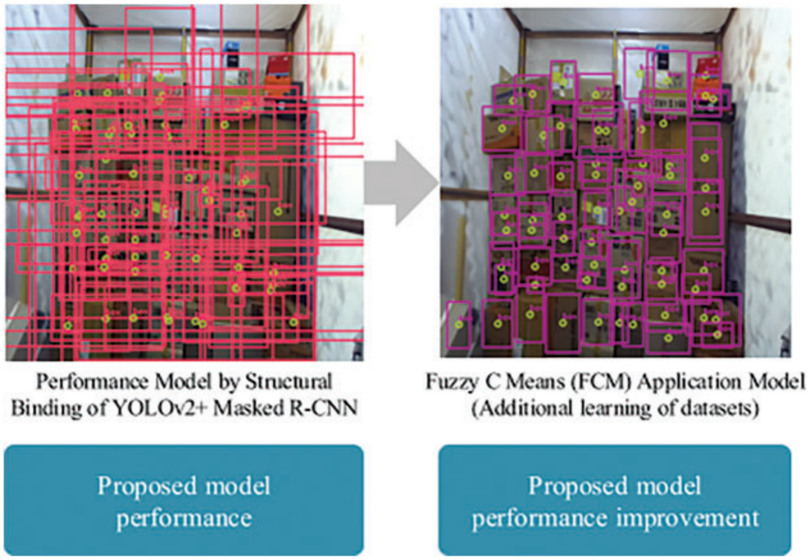
Performance Model by Structural Binding of YOLOv2+ Masked R-CNN
Fuzzy C Means (FCM) Application Model (Additional learning of datasets)
Proposed model performance
Proposed model performance improvement

Figure 9: Composition of fuzzy convergence algorithm

5 Conclusion

This study was proposed based on the YOLOv2 model, which enables rapid object detection and classification for cargo recognition of automatic unloading equipment. Mask R-CNN, which enables the detection of accurate pixel units of cargo with YOLOv2, and Fuzzy C Means were applied to improve recognition and performance. To verify the performance of the fuzzy convergence algorithm, the experiment was conducted by building a space similar to the actual work space. As a result, the algorithm proposed through this study showed excellent performance in cargo object recognition rate and real-time detection rate. Satisfactory performance results are expected to produce the same results in real-world environments. This study proposed an algorithm to enhance the recognition performance of unstructured courier cargo by selecting an automatic up-and-down unloading device in the logistics field. The fuzzy convergence algorithm is a proposed form combining the existing algorithms YOLOv2 and Mask R-CNN. When comparing each algorithm, the fuzzy C mean convergence algorithm has a lower FPS mean than YOLOv2 and a lower recognition rate than Mask R-CNN. However, the fuzzy C mean convergence algorithm compensated for the shortcomings of the two existing algorithms and showed high cargo recognition bounding box density of the existing algorithms. Furthermore, the fuzzy convergence algorithm was applied with the fuzzy C means method, which improved the accuracy of recognizing objects in more detail. The results of this study can be utilized in efficient automatic cargo unloading systems and, more broadly, have a positive impact on the logistics sector. The object recognition field is showing progress from day to day, and in the future, research will be conducted to improve algorithms to meet this trend.

Funding Statement: This work was supported by a grant from R&D program of the Korea Evaluation Institute of Industrial Technology (20015047).

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

References

[1] H. J. Shin, "A study on trends in the use of logistics technology based on the 4th industrial revolution," *The e-Business Studies*, vol. 21, no. 2, pp. 17–27, 2020.

[2] K. J. Kwak, S. Y. Hwang, D. J. Shin, K. G. Park, J. J. Kim *et al.,* "Study of logistics object tracking service for smart SCM," *Journal of the Korean Institute of Industrial, Engineers*, vol. 46, no. 1, pp. 71–81, 2020.

[3] W. P. Yu, Y. C. Lee and D. H. Kim, "Technical trends of robot task intelligence in intelligent logistics/agriculture," *Electronics and Telecommunications Trends*, vol. 36, no. 2, pp. 22–31, 2021.

[4] S. l. Choi, D. H. Kim, J. Y. Lee, S. H. Park, B. S. Seo *et al.,* "Logistics and delivery robots in the 4th industrial revolution," *Electronics and Telecommunications Trends*, vol. 34, no. 4, pp. 98–107, 2019.

[5] C. H. Park, S. D. Bae, S. G. Choi, S. H. Choi, J. W. Choi *et al.,* "Automatic picking/classification system using video analysis," in *Proc. of the Korean Society of Computer Information Conf.*, Jeju, Korea, vol. 28, no. 2, pp. 661–662, 2020.

[6] J. U. Won, M. H. Park, S. W. Park, J. H. Cho and Y. T. Kim, "Deep learning based cargo recognition algorithm for automatic cargo unloading system," *Journal of Korean Institute of Intelligent Systems*, vol. 29, no. 6, pp. 430–436, 2019.

[7] H. C. Hwang and S. H. Song, "A study on the factors affecting the acceptance of logistics robot in the fulfillment center using the technology acceptance model," *Journal of the Korea Academia-Industrial Cooperation Society*, vol. 20, no. 12, pp. 287–297, 2019.

[8] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[9] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation," D. E. Rumelhart, J. L. McClelland and the PDP Research Group, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, Cambridge, MA, USA: MIT Press, pp. 318–362, 1986.

[10] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005.

[11] J. K. Park and J. B. Park, "An object recognition method based on depth information for an indoor mobile robot," *Journal of Institute of Control, Robotics and Systems*, vol. 21, no. 10, pp. 958–964, 2015.

[12] S. I. Park and M. Y. Kim, "Multiple-background model-based object detection for fixed-embedded surveillance system," *Journal of Institute of Control, Robotics and Systems*, vol. 21, no. 11, pp. 989–995, 2015.

[13] H. Bay, T. Tuytelaars and L. V. Gool, "Surf: Speeded up robust features," in *Computer Vision–European Conference on Computer Vision*, Graz, Austria, pp. 404–417, 2006.

[14] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the 7thIEEE Int. Conf.*, Kerkyra, Greece, vol. 2, pp. 1150–1157, 1999.

[15] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[16] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580–587, 2014.

[17] Q. Chen, X. Gan, W. Huang, J. Feng and H. Shim "Road damage detection and classification using mask R-cNN with denseNet backbone," *Computers, Materials & Continua*, vol. 65, no. 3, pp. 2201–2215, 2020.

[18] R. Meng, Q. Cui, Z. Zhou, C. Yuan and X. Sun "A novel steganography algorithm based on instance segmentation," *Computers, Materials & Continua*, vol. 63, no. 1, pp. 183–196, 2020.

[19] C. B. Murthy, M. F. Hashmi, G. Muhammad and S. A. AlQahtani, "YOLOv2PD: An efficient pedestrian detection algorithm using improved YOLOv2 model," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3015–3031, 2021.

[20] J. Amin, M. A. Anjum, M. Sharif, S. Kadry, Y. Nam *et al.,* "Convolutional Bi-lSTM based human gait recognition using video sequences," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 2693–2709, 2021.

[21] J. Amin, M. Sharif, M. A. Anjum, A. Siddiqa, S. Kadry *et al.,* "3D semantic deep learning networks for leukemia detection," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 785–79, 2021.

[22] Q. Zhou, J. Qin, X. Xiang, Y. Tan and N. N. Xiong, "Algorithm of helmet wearing detection based on at-yOLO deep mode," *Computers, Materials & Continua*, vol. 69, no. 1, pp pp. 161–174, 2021.

[23] A. Baccouche, B. G. Zapirain, C. C. Olea and A. S. Elmaghraby, "Breast lesions detection and classification via YOLO-based fusion models," *Computers, Materials & Continua*, vol. 69, no. 1, pp pp. 1407–1425, 2021.

[24] B. Matija, °P. Miran and °M. I. Kos, "Ball detection using YOLO and mask R-cNN," in *Conf.: 2018 Int. Conf. on Computational Science and Computational Intelligence*, Las Vegas, NV, USA, pp. 319–323, 2018.

[25] M. A. Soeleman, M. Hariadi and M. H. Purnomo, "Adaptive threshold for background subtraction in moving object detection using fuzzy C means clustering," in *TENCON IEEE Region 10 Conf.*, Cebu, Philippines, 2012.

[26] E. S. Kim and K. S. Joo, "The pattern segmentation of 3D image information using FCM," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 10, no. 5, pp. 871–876, 2006.