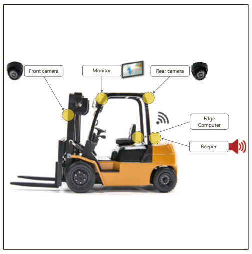

Edge AI-based Forklift Safety Support System

Shinya TAKUBO*, Atsuki NAGATA, Shungo IKAI, Mofei LI,
Shingo NISHIOKA, and Akinobu YOSHIMURA

To mitigate accidents involving forklifts and people, we have developed a forklift safety support system capable of detecting the presence of people near a forklift. Real-time responsiveness is crucial, necessitating immediate detection and reporting. To address this, we propose utilizing edge computing to enable swift responses. This paper introduces a forklift safety support system leveraging edge AI installed in a compact edge computer. Additionally, we discuss the evaluation of edge AI's accuracy and speed at the edge, along with the examination of techniques to enhance the accuracy efficiently.

Keywords: forklift safety support, edge AI, training tool, quantization, FPGA

1. Introduction

Sumitomo Electric Industries, Ltd. has developed a worker detection system as a means of forklift safety support to improve the manufacturing factory. The purpose of this system is to report to the forklift driver about the approach of a worker(s) towards the forklift and to support production site safety by reviewing the site condition at a later time. Techniques used to realize this system include camera-, ultrasound-, and IC tag-based systems. Sumitomo Electric has employed a camera-based system (Table 1).

It is desirable that the process from worker detection to reporting takes place in real time. Sumitomo Electric has developed a small edge computer (Table 2), which is installed on each forklift and used in the judgment process. Moreover, it is necessary to detect workers accurately, independently of helmets, work clothes, and height, or whether they are carrying an object or not, that is, the varying patterns conceivable in a factory. To meet this task, a deep neural network (DNN),*[1] an important and evolving part of artificial intelligence (AI) systems in recent years, is installed on the FPGA*[2] in the edge computer to achieve both satisfactory judgment accuracy and speed.

Chapter 2 of this paper explains the forklift system configuration and the overall configuration of an edge-and-cloud system, including the DNN development cycle; Chapter 3 describes the evaluation of the edge AI provided to achieve both satisfactory judgment accuracy and speed; Chapter 4 explains an efficient model development technique for worker detection.

Table 1. Comparison of Nearby Worker Detection Systems

System	UltraSound	IC tag	Camera
Object detected	Objects reflecting ultrasound	IC tag	People
Marker	not required	required	not required
Detection range	~4 m	~10 m	~10 m
Video image display	N/A	N/A	available
Video recording	N/A	N/A	available

Table 2. Edge Computer Specifications

Dimensions [mm]	115 mm × 63 mm (W×H)
Interface with peripheral devices	USB, Display Port, GPIO, and so on
Processor	Arm+FPGA
OS	Linux

2. System Configuration

The configuration of the system is presented in Figs. 1 and 2.

Figure 1 shows the forklift system configuration. One camera is installed in the front and another in the rear; these cameras are connected to the edge computer installed close to the driver's seat. When a nearby worker is detected in an image, the system displays the detection situation on the monitor installed above the driver's seat and sounds a beep. The edge computer, connected to the internal wireless network, is capable of sending images and detection results to a cloud server. It is also possible to monitor the operating status via the cloud.

Fig. 1. Forklift system configuration

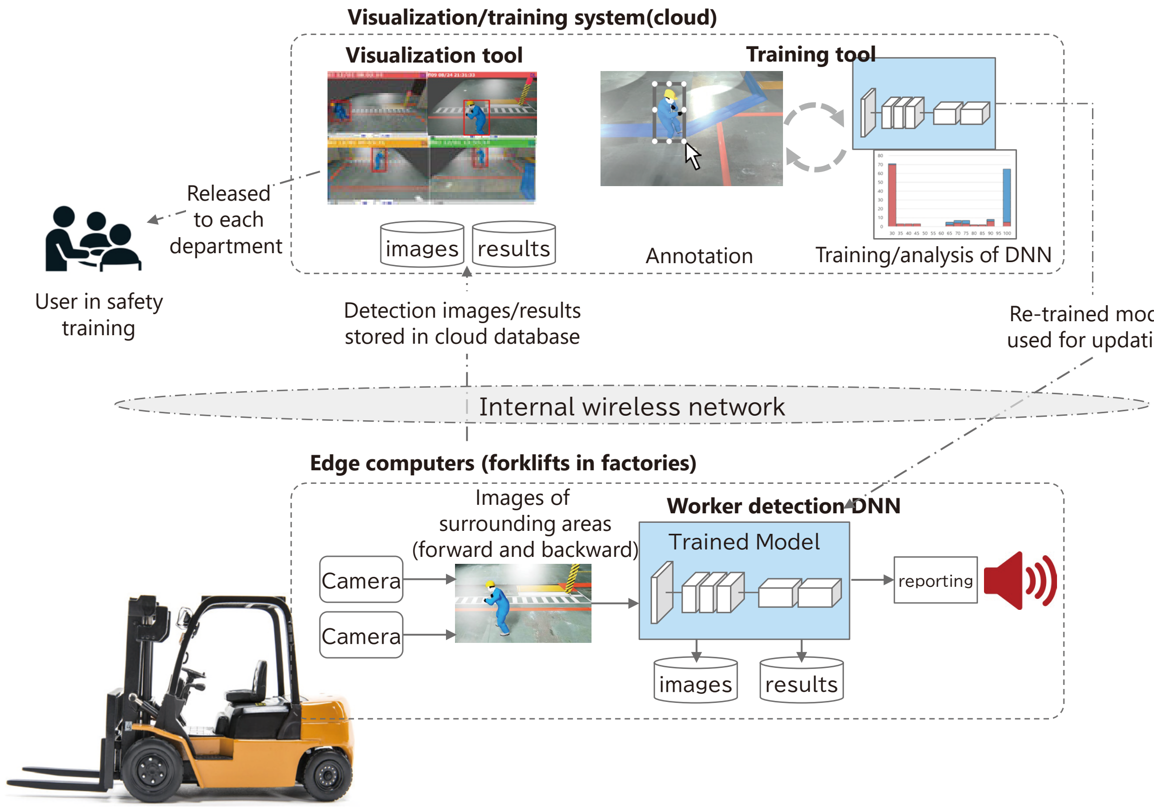

Fig. 2. Overall configuration of edge & cloud system

Figure 2 illustrates the overall configuration including cloud-side functions. When the DNN installed on the edge computer detects a nearby worker, the images and results are saved in the storage in the edge computer and sent to a cloud database at the same time via the internal wireless network. There is storage on the cloud side, which can store data on a yearly basis. Therefore, videos from specific periods can be reviewed later. Moreover, a visualization tool, which analyzes risks based on the images and results, has been released to each department for use in safety training.

Automated data acquisition eliminates the need to search through massive amounts of data for objects to be detected and enables efficient annotation,*3 as shown in Fig. 2.

Moreover, while DNN training normally requires advanced programming skills and building a complex environment, a tool that enables a training cycle to be followed by simply using a web browser and a mouse has been developed and utilized (Fig. 3 (a)).

The DNN training results are quantified and managed by a history and accuracy chart (confusion matrix; Fig. 3 (b)). Moreover, to find error trends by images, a function is installed to allow the user to jump to the corresponding image by clicking on a desired cell in the accuracy chart, which enables the weak points of the DNN to be analyzed bilaterally from quantified values and images.

The trained DNN is used to remotely update edge computers via the internal wireless network. In this way, DNNs are operated and managed at both domestic and overseas sites.

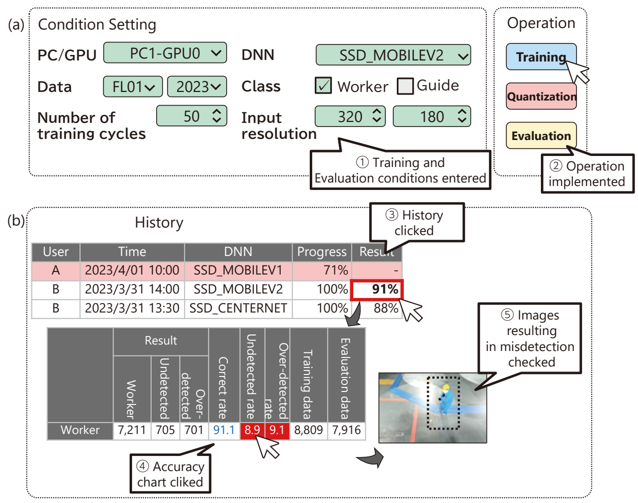

Fig. 3. Independently developed DNN training tool

3. Development of Edge AI

Deep neural networks are known to deliver higher performance than image processing and other conventional systems by training various data at the cost of a huge amount of computation. Therefore, in recent years, research has been promoted on edge AI to implement a DNN while ensuring fast and accurate operation using a small device with limited resources. To realize edge AI, we conducted: ① benchmarking of DNNs suitable for edge computing and ② evaluation of quantization bits intended to achieve downsizing while mitigating accuracy degradation. Facts about the evaluation described in ¬this chapter are given in Table 3.

Table 3. Evaluation Factors

Data	Worker data acquired in Sumitomo Electric's factories has reached approximately 15,000 images
Judgement threshold	Reliability threshold : 0.3 IoU threshold : 0.1

The reliability threshold is a threshold for determining whether or not to make an output according to the reliability assigned to detection boxes (described later in 4-1). The IoU threshold is a threshold for determining whether or not to make an output according to the overlap between the correct solution and a detection box.

3-1 Benchmarking of DNNs suitable for edge computing

Deep neural networks are grouped into several types by task. Because the current focus is on an application that detects nearby workers, an object detection task DNN was selected, which determines the locations of objects in images. Among object detection DNNs, Single Shot Multibox Detector (SSD) is known to be fast and accurate.(1) Many methods derived from this method have been proposed. Their accuracy depends on training data sets. In the current study, DNNs suitable for edge computing were benchmarked using Sumitomo Electric's image data containing nearby workers around forklifts. Note that the accuracy values presented in this paper have all been determined on a worker box-by-worker box basis.

Figure 4 presents the results of the comparisons. The number of arithmetic operations is plotted on the x-axis; accuracy is plotted on the y-axis. The bubble size represents the parameter size. Small bubbles in the upper-left area of the graph are desirable in terms of accuracy, speed, and size. In the current evaluation, eight types of DNN were tested. While SSD_VGG16 and SSD_MOBILENETV3_SMALL are advantageous in terms of accuracy and speed, respectively, SSD_MOBILENETV2 was determined to be suitable in terms of balance.

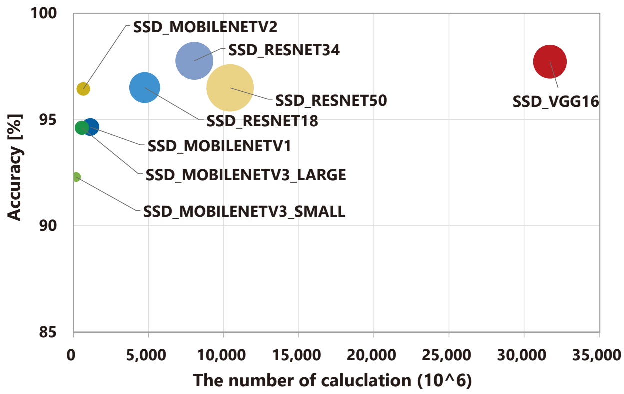

Fig. 4. DNN benchmarking

3-2 Evaluation of quantization bits

The DNN repeats a large amount of product-sum operations of parameters and input values, as illustrated in Fig. 5. In this process, the quantization technique is available, which is intended to reduce the size of the arithmetic circuit by expressing each operation with the 8-bit integer (INT8) type or the like instead of 32-bit floating-point accuracy (float32). For example, a comparison of 32-bit and 8-bit computing shows that the amount of 8-bit computing is 1/16 of that of 32-bit computing [(8 × 8) ÷ (32 × 32)]. Thus, quantization is essential for realizing edge AI in terms of faster operation and power conservation. In the present study, the accuracy and speed achieved by quantization sequentially down to 1 bit were verified against the standard 8-bit computing commonly used in edge AI. The technique used in this paper is known as post-training quantization (PTQ), which carries out DNN quantization after training.

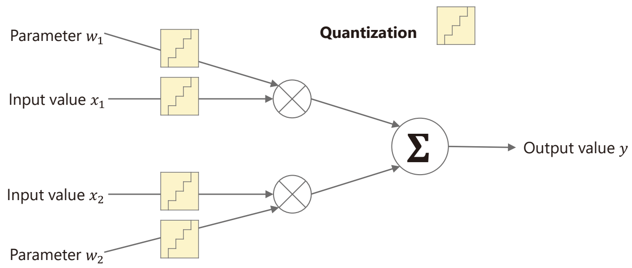

Fig. 5. Principle of quantization

Figure 6 presents a graph, which plots the number of quantization bits (a decreasing order of parameter × input value) on the x-axis and accuracy [%] on the y-axis. In the evaluation, quantization bits varied from 8 bits to 1 bit. No difference in accuracy was observed between before and after quantization down to 6 bits, while quantization down to 5 bits and below of parameters and input values resulted in a noticeable decrease in accuracy.

Next, the study focused on what impact parameters and input values would have on quantization. Figure 7, which plots the number of quantization bits on the x-axis and accuracy [%] on the y-axis, shows the results of accuracy comparison in which either the parameter or input value was fixed at 8 bits and the other varied from 8 bits to 1 bit. According to the results, accuracy decreased noticeably when the parameter or input value was quantized down to 5 bits or below, with no difference in impact being observed. The amount of computation for 6 bits × 6 bits is larger than for 8 bits × 5 bits (or the reverse order); however, 6 bits × 6 bits resulted in better accuracy. Consequently, the quantization bit balance between parameters and input values, as well as the simple amount of computation, was found to have an impact on accuracy.

The results revealed that to establish a good balance between accuracy and speed with the edge AI of the current system, assigning 6 bits for the parameter and 6 bits for the input value is suitable as quantization bits, and that the amount of computation can be reduced to about 1/28 compared to that before quantization and to about 2/3 compared to common 8-bit quantization.

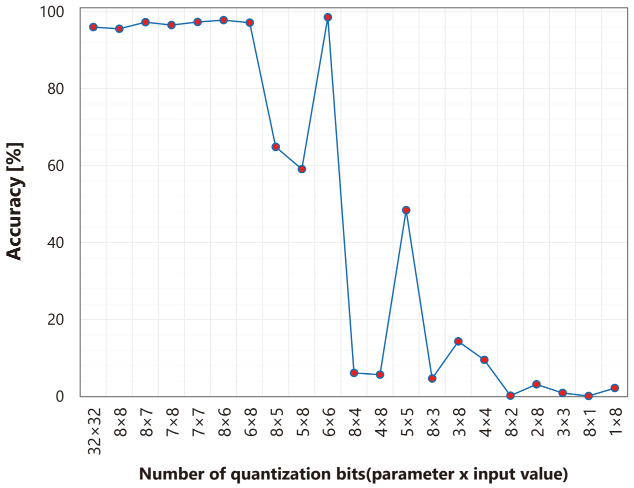

Fig. 6. Accuracy comparison with respect to different numbers of quantization bits

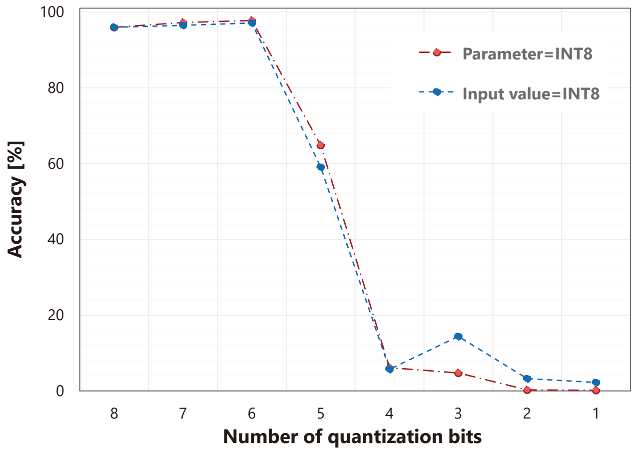

Fig. 7. Accuracy comparison in which either the parameter or input value fixed at 8 bits

4. Development of Worker Detection DNN

To improve DNN accuracy, the number of training data is important. This system automated the data collection process by sending data to the cloud as we showed in Chapter 2. However, human annotation is still required. In order to improve DNN accuracy more efficiently, we verified a data augmentation technique that automatically generates training data from the false positive detections.

4-1 Problem of worker detection DNN

Figure 8 shows a DNN trained by 10,000 data collected from the factory, it erroneously detected a forklift approach light as a worker. This is because approximately 90% of the workers wore blue working clothes in training data, so the DNN tended to mistake a blue object as a worker. It is possible to solve this problem by defining the approach light as a new class and training a 2-class object detection DNN. However, It takes time to collect approach light data and annotate them. We developed an approach to increase training data efficiently by combining the approach light data with background images.

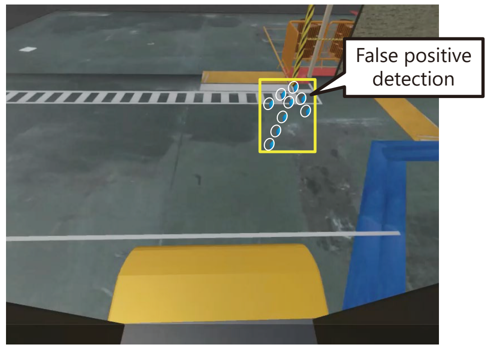

Fig. 8. False positive detection of approach light

4-2 Data augmentation

Figure 9 shows the approach of data augmentation of training data. The procedure 1-3 is listed below.

1. Divide all data into a set of images (A) which has an approach light and the other images (B) which is background only.
2. Crop the position of approach light in (A) and extract the blue part from the cropped image.
3. Attach the blue part to a background image from (B) and get synthetic set of images (C).

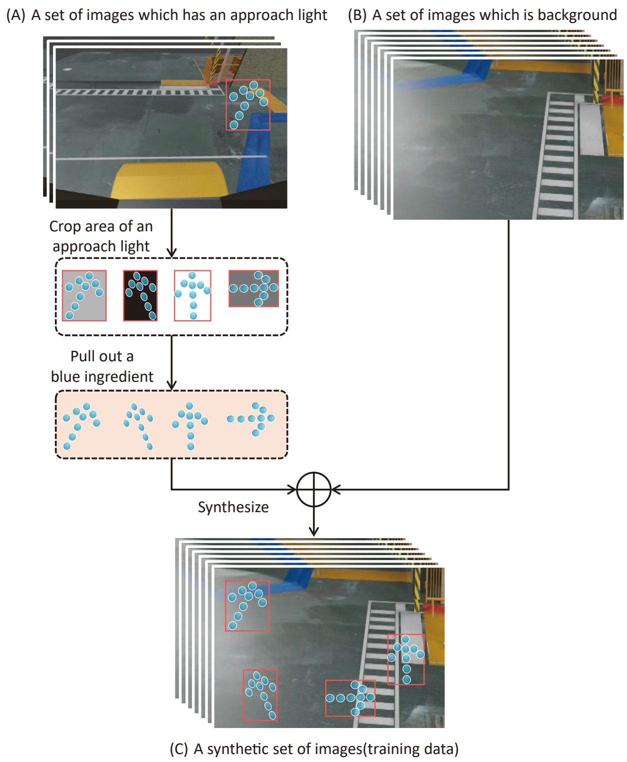

Fig. 9. The approach of data augmentation

The maximum number of approach lights attached is five, and the position is placed on the image randomly. In addition, by saving the position of the approach light simultaneously, the training data is automatically created and annotation is unnecessary.

4-3 Experimental Results

Figure 10 shows the results. In this experiment, 10,000 data are trained including 439 data with approach light. We compared the accuracy of training with only 10,000 real data and training with the addition of 10,000, 20,000, and 30,000 synthetic data. In consideration of randomness of the training, we trained five times on each condition.

Figure 10 shows that DNN trained with 10,000-20,000 synthetic data can improve accuracy by 2-5% in comparison with DNN trained only real data. It shows that by adding a new approach light class to DNN, DNN can distinguish worker and the difference of the approach light. However, when trained with too many synthetic data (over 30,000), the accuracy was not improved. It shows that the balance between the real data and synthetic data is important (In this case the ratio of real/synthetic data is 1:3).

While creating synthetic data, the training data is automatically created. So, this approach can increase training data very efficiently, and improve accuracy effectively.

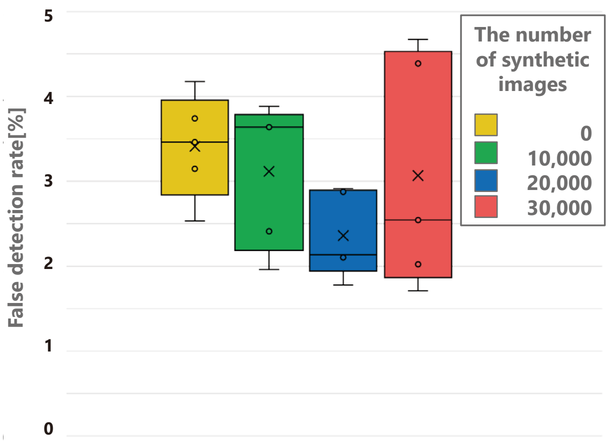

Fig. 10. Accuracy comparison based on the number of synthetic images

5. Conclusion

A forklift safety support system was developed, which performs detection and reporting in real time with an edge computer and edge AI mounted on a forklift. Benchmark results of the AI for edge computing that achieves both satisfactory accuracy and speed and training approach to efficiently improve accuracy were described. Going forward, we intend to further improve the accuracy and speed of the DNN, while continuing to prevent accidents and ensure the safety of the factory.

Technical Terms

*1 Deep Neural Network: A type of artificial intelligence (AI). This technique is used to deliver outputs from input data through repeated multiple product-sum operations, in which parameters used in arithmetic operation are automatically adjusted by advance training. Compared to conventional neural networks, which use one layer of intermediate neural network operations, the deep neural network consists of two or more intermediate layers and can accommodate complex data processing.

*2 FPGA: Abbreviation for field programmable gate array. The FPGA is a circuit that allows the user to rewrite the logic and wiring in a flexible manner. Compared to graphics processing units (GPUs) in general use as AI accelerators, the FPGA offers high operation customizability and, when combined with quantization technology, can realize highly power-efficient edge AI.

*3 Annotation: The process of creating teaching data during supervised learning, which is a type of machine learning, by assigning labels and other metadata to data.

Reference

(1) Wei Liu, "SSD: Single Shot MultiBox Detector," European Conference on Computer Vision 2016, pp.21-37

Contributors

The lead author is indicated by an asterisk (*).

S. TAKUBO*
• IoT R&D Center

A. NAGATA
• IoT R&D Center

S. IKAI
• IoT R&D Center

M. LI
• IoT R&D Center

S. NISHIOKA
• IoT R&D Center

A. YOSHIMURA
• Group Manager, IoT R&D Center