



Development of an AI-Equipped 5G Terminal and Demonstration of Automatic Performance Improvement in Pedestrian Detection Sensors

Katsuhiko YADA*, Shunsuke MAKI, Makitaro FURUTA, Shunsuke KADOI, Takuya SUZUKA, and Shohei OGAWA

Our company has been shipping the industrial 5G terminal IGW5111. As demand for performing AI functions through edge processing is expected to increase, we developed an AI-equipped 5G terminal capable of supporting such functions. As part of the Cross-ministerial Strategic Innovation Promotion Program (SIP) Phase 3/Development of Smart Mobility Platform, we aimed to automatically improve the performance of image detection AI used for pedestrian detection at intersections. Through demonstration experiments, we verified the effectiveness of the algorithm and edge processing. Furthermore, we verified the usefulness of edge processing in improving safety in factories.

Keywords: 5G terminal, edge processing, object detection AI, annotation

1. Introduction

Along with the development of wireless communication technology, communication protocols used in mobile phones have been advancing. Currently, 5G is the widely used communication system. The 5G communication protocol provides high data rates, great capacity, and low latency, and is being increasingly adopted in industrial applications, as well as for consumer uses. Sumitomo Electric Industries, Ltd. has developed an artificial intelligence (AI)-equipped 5G terminal that can incorporate AI capabilities and has already been shipping the industrial-use 5G terminal IGW5111 on a commercial basis. This report outlines this terminal and describes examples of its application to road traffic management (automatic performance improvements of pedestrian detection sensors) and factory safety improvements.

2. Development of an AI-Equipped 5G Terminal

Wireless LAN and other wireless communication systems often experience unstable connections caused by interference and other factors. In contrast, 5G communications use a licensed band. Moreover, by using a dedicated frequency, 5G for industrial uses (local 5G) achieves very stable communications. Local 5G, however, requires a license for the frequency used, a base station set up on one's own, and a large amount of capital investment. As a solution, SoftBank Corporation (hereinafter referred to as "SoftBank") offers a stable communication environment named Private 5G for industrial applications, having built private 5G networks, which utilize frequencies allotted to SoftBank, for exclusive use by its customers. Users are freed from the need to set up a base station and can reduce initial investment. Sumitomo Electric has developed and commercialized the industrial 5G terminal IGW5111 for Private 5G and has been providing it to SoftBank.

The IGW5111 is equipped with 5G communication and wireless LAN capabilities, among other functions, and

comes with the Docker functionality installed, which runs user applications on the main unit. Consequently, the IGW5111 performs edge processing without any external devices. However, due to limitations of the main unit's CPU, it cannot perform AI functions such as deep learning. Nowadays, AI has come into widespread use at a remarkable pace. There is a strong demand for AI-powered edge computing in terms of the following factors: (1) reduced data traffic, (2) low latency, and (3) uninterrupted provision of services in the event of a communications blackout. Therefore, we have developed an AI-equipped 5G terminal, which is an IGW5111 incorporating a graphics processing unit (GPU) capable of performing AI functions. Table 1 shows the specifications for the AI-equipped 5G terminal.

Table 1. Specifications for the AI-equipped 5G terminal

Features	Specifications
5G	3GPP Rel-16-compliant, stand-alone (SA) mode Supported frequency: Sub-6 GHz and millimeter waves
Wireless LAN	IEEE802.11 a/b/g/n/ac/ax
Wired LAN	100/1000BASE-T × 3
Other I/F	RS-485 × 1, contacts (DI × 1, DO × 1)
Size	315 × 300 × 82 mm
Weight	4.0 kg
Cooling system	Self-cooling (fanless)
AI capabilities	NVIDIA Jetson Orin NX, 16 GB 1024 NVIDIA CUDA cores, 157 TOPS

3. Automated Improvement of the Performance of Pedestrian Detection Sensors

3-1 Objective

As the Vision Zero concept is being promoted in many countries in order to eliminate deaths and serious injuries caused by traffic accidents, development of road traffic management systems is thriving with the goal of ensuring the safety of vulnerable road users, including

pedestrians and bicycles. Japan as well has the Cross-ministerial Strategic Innovation Promotion Program (SIP) Phase 3/Development of Smart Mobility Platform, in which “research and development of support for the prevention of traffic accidents by means of prior risk notification” are promoted. As a member of this project, Sumitomo Electric has worked on developing a pedestrian detection sensor and conducting demonstration experiments in an actual environment with the goal of preventing accidents resulting from pedestrians abruptly emerging at intersections with blind spots.

Broadly speaking, the capability of image recognition AI to detect objects is not very high in the actual field. This is because object recognition AI has been adjusted for general purposes to detect all kinds of images. For pedestrian detection applications, however, the field of view is fixed. So, it is nonetheless possible to significantly improve AI’s object detection capability by creating training data from images collected on site and conducting additional training. That said, for improved detection capability, it is necessary to acquire thousands of images on site where pedestrian detection sensors are installed, manually specify the positions and categories of objects (annotation^{*1}), and conduct additional training using the data as training data.

Suppose, for example, that 3,000 images are taken per sensor installation location for the purpose of additional training and annotation requires approximately 40 seconds per image. Then, it takes 33 hours and 20 minutes for annotation alone. This task must be carried out for each sensor installation location. Consequently, sensors installed at 100 locations, for example, require a total of 3,333 hours. The cost of doing so is not included in the price of the main unit or construction costs. Moreover, it becomes necessary to provide skilled operators to be assigned to annotation, thereby posing a significant challenge to social implementation.

To overcome this challenge, this project aims to reduce social implementation costs by developing a function that automatically performs annotation and creates training data.

3-2 Overview of automatic annotation

Figure 1 outlines a configuration which applies automatic annotation to pedestrian detection.

The pedestrian detection sensor is equipped with a fisheye lens, an object detection AI model (You Only Look Once [YOLO]), and pedestrian detection. Installed at an intersection, it captures images at a frequency of every 100 ms and determines the positions of pedestrians and bicycles as well as their possibilities of crossing the road at a cross-

walk or other places.

The AI-equipped 5G terminal sends to the traffic surveillance center and other recipients the positions of pedestrians and bicycles as well as their possibilities of crossing the road detected by pedestrian detection sensors. Powered by an object detection AI model (Collaborative-Detection Transformer [Co-DETR]), which differs from that on the pedestrian detection sensor, the AI-equipped 5G terminal detects pedestrians and bicycles independently from the pedestrian detection sensor. Annotation (generation of bounding box^{*2}-category combinations) is done by comparing both results. Generated training data is transmitted to the training server as data for additional training. Although automatic annotation and generation can be implemented on a training server, edge processing is carried out to reduce data traffic between the AI-equipped 5G terminal and the training server.

The traffic surveillance center receives information as to whether or not any pedestrian or bicycle with high probability of crossing the road is present and notifies and alerts drivers of nearby moving automobiles and other vehicles.

The training server accumulates training data, updates YOLO and Co-DETR training models, and feeds the results back to the object detection AI on the pedestrian detection sensor and the AI-equipped 5G terminal.

The Co-DETR equipped in the AI-equipped 5G terminal has higher detection capability than that of YOLO because it devotes more time to processing than YOLO does. Even so, it is subject to detection errors. In particular, its detection performance is unsatisfactory when it has not been trained on images from sensor installation locations. For this reason, training data is created individually for YOLO and Co-DETR for their respective training sessions and feedback (red-shaded and blue-shaded areas in Fig. 1).

3-3 Automatic annotation

Both YOLO and Co-DETR compute confidence^{*3} for each object when they detect them. This, however, is unsuitable for the generation of annotation data because detection errors occur frequently despite high confidence. Since YOLO generates object bounding boxes every 100 ms, we decided to perform annotation by tracking the positions of objects. More specifically, the positions and categories of objects are determined on a majority rule basis using three pieces of information: (1) detection results of YOLO, (2) positions and categories inferred from preceding and following images by means of object tracking, and (3) Co-DETR results. Table 2 gives examples of inference result criteria and recipients.

The principal features of this system are described below.

- (1) The fisheye camera on the pedestrian detection sensor captures intersection images every 100 ms.
- (2) The object detection model (YOLO) equipped on the pedestrian detection sensor makes inferences.
- (3) The system tracks YOLO bounding boxes.
- (4) Because the processing speed of Co-DETR is slow, images inferred by YOLO are thinned out.
- (5) Using images captured by the pedestrian detection sensor and thinned out (due to Co-DETR’s slow processing), Co-DETR detects objects.
- (6) The system performs annotation on a majority rule basis (see Table 2) using the results produced by YOLO

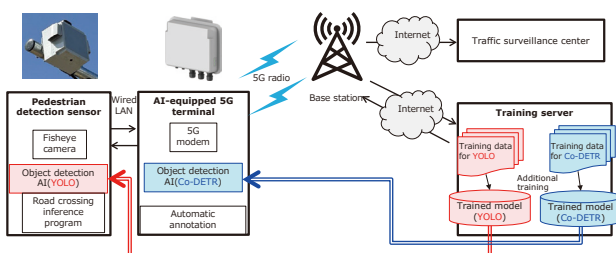


Fig. 1. Automatic annotation system configuration

and Co-DETR. Those which show little motion as a result of object tracking are disregarded as still objects.

(7) The annotation results are sent to the training server as training data.

(8) When a predetermined amount of training data has been accumulated, the training server conducts additional training to create new trained AI models.

(9) The trained AI models are applied to the object detection AI of the pedestrian detection sensor and AI-equipped 5G terminal for updating.

With the processes from the aforementioned capture of images and AI model application and updating being one cycle, improvement of the object detection performance is pursued through repeated cycles.

Table 2. Examples of inference result criteria and recipients

YOLO result	Object tracking	Co-DETR result	Motion detection	Annotation (output)	
				Category	Recipient
Person	Person	Person	-	Person	Both
Person	Bicycle	Person	-	Person	Both
Person	Person	Bicycle	-	Person	Co-DETR
Person	Bicycle	Bicycle	-	Bicycle	YOLO
Person	Person	No detection	-	Person	Co-DETR
Person	No detection	No detection	-	No detection	YOLO
No detection	Person	Person	-	Person	YOLO
No detection	No detection	Person	Detected	Person	YOLO
Other combinations				Unable to determine	Out of scope

3-4 On-public-road demonstration experiment

The operation of the automatic annotation was tested on a public road by installing the pedestrian detection sensor and AI-equipped 5G terminal at an intersection in Tsukuba City during the period from September 24, 2025 to March 24, 2026. The installation location used at that time is illustrated in Fig. 2, how they were installed is pictured in Photo 1, and an actual image captured by the pedestrian detection sensor is given in Photo 2.

This demonstration experiment aimed to verify automatic performance improvements in the pedestrian detection sensor installed in the field. Therefore, the detection categories were limited to person, bicycle, and umbrella. Meanwhile, this demonstration experiment restricted automatic annotation to the period between sunrise and sunset because there were no street lights in the vicinity of the intended intersection and the camera on the pedestrian detection sensor was not designed for night photography.

The following rules were applied when compiling object detection performance data.

- (1) Exclude bicycles parked in bicycle parking areas from detection because the goal of the object detection is to infer the presence of crossing pedestrians. (When any such bicycle is detected, classify it as over-detection).
- (2) Place each person walking a bicycle in the same bounding box as the bicycle and categorize him or her as bicycle. (Classify separate detection of the person and bicycle as over-detection.)

- (3) When a person is in a bicycle parking area, detection either as person or bicycle is correct because doing so has no influence on the inference regarding crossing the road.
- (4) Detection of a person under umbrella either as person or umbrella is correct. However, classify detection of both as over-detection.



Fig. 2. Installation location in on-public-road demonstration experiment (marked with a circle on the map)



Photo 1. Installation status of devices during the on-public-road demonstration experiment



Photo 2. Example image taken by the pedestrian detection sensor

- (5) Avoid counting babies or infants held in the arms of people.
 (6) Classify the detection of a motorcycle or scooter as over-detection because they are out of the scope of training.

3-5 Accuracy of automatically generated annotation data

To ascertain the performance of automatic annotation at the actual intersection, the annotation results produced at the intersection on site were manually evaluated and compiled as to their validity. Table 3 lists accuracy, detection rate, precision,*⁴ recall,*⁵ and F-measure, with the annotation results interpreted as the results produced by the object detection AI.

Of these indicators, the accuracy is the percentage of cases where objects were correctly detected in correct categories, while the detection rate refers to the percentage that objects were correctly detected while overlooking category errors.

Regarding the annotation results that were incorrect in cycle 1 (used for training cycle 2), the factor classification results are presented in Table 4.

Table 3. Automatic annotation performance evaluation

Trained model used		Initial model	Cycle-1
Accuracy		95.8%	97.5%
Detection rate		97.5%	98.6%
Precision	Person	98.0%	99.6%
	Bicycle	96.2%	97.8%
Recall	Person	96.8%	96.3%
	Bicycle	95.7%	99.4%
F-measure	Person	97.4%	98.0%
	Bicycle	96.0%	98.6%

Table 4. Factor classification results for incorrect annotation results

Factor		Percentage in all incorrect results	
Undetected	Person in bicycle parking area	55.77%	38.46%
	Overlap with another person		5.77%
	Infant with a parent		3.85%
	Infant in baby stroller		3.85%
Over-detection	Motorcycle detected as bicycle	1.92%	1.92%
Category error	Person walking a bicycle detected as person	42.31%	1.92%
	Person walking a shopping cart or the like detected as bicycle		19.23%
	Other (bicycle → person)		5.77%
	Other (person → bicycle)		15.38%

Missed detection of people in bicycle parking areas accounted for 38.46% of incorrect-detection patterns. An example is given in Photo 3 (detection missed in the red-encircled area). People in bicycle parking areas often show little motion as they put their baggage on their bicycles or lock their bicycles. The current object tracking was designed to remove objects from annotation as still objects if they are motionless for a long time. This is likely the cause of missed detections. Improvements can probably be

made by adjusting the threshold that determines still objects. Another possible cause is that pre-training lacked persons with bicycles in the background, so improvements can also be possibly made by improving pre-training.

Meanwhile, the pattern in that people walking a baby stroller or shopping cart were erroneously categorized as bicycle accounted for 19.23% of incorrect detections. One possible cause of this is that baby strollers, shopping carts, and the like were not included in the detection categories and were not used for training. This issue is expected to improve through pre-training given before installation in the field, providing training using separate categories for baby strollers, shopping carts, and the like.

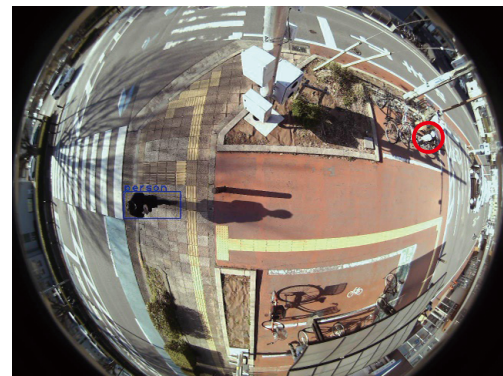


Photo 3. Undetected person in a bicycle parking area

3-6 Detection performance improvement effects on pedestrian detection sensors

When ascertaining the performance of an object detection AI model, it is necessary to evaluate each cycle under identical conditions. Therefore, 3,695 images were selected in advance for evaluation from those picturing the intersection on site. Annotation was performed manually using the selected images to create correct data. After detection processes by each trained model on the evaluation-use images, the detection results were compared with the correct data. Table 5 shows the comparison results, which reveal that one-cycle training of the current algorithm improved the accuracy from 52.7% to 91.8% and the detection rate from 55.7% to 93.9%.

Table 5. Evaluation of performance of each trained model

Trained model used		Initial model	Cycle-1	Cycle-2
Accuracy		52.7%	91.8%	92.2%
Detection rate		55.7%	93.9%	94.2%
Precision	Person	62.7%	94.6%	97.0%
	Bicycle	40.1%	93.5%	93.1%
Recall	Person	93.4%	94.5%	92.5%
	Bicycle	86.6%	95.2%	98.1%
F-measure	Person	75.0%	94.6%	94.7%
	Bicycle	54.8%	94.4%	95.6%

Additionally, regarding the incorrect inference results, detailed factor analysis results for cycle 2 are presented in Table 6.

Table 6. Analysis results for factors contributing to incorrect YOLO inference results

Factor		Percentage in all incorrect results	
Undetected	Person in bicycle parking area	47.47%	17.98%
	Overlap with another person		6.87%
	Infant with a parent		6.26%
	Infant in baby stroller		8.89%
	Other		7.47%
Over-detection	Motorcycle → Bicycle	27.47%	3.84%
	Bicycle in bicycle parking area		1.41%
	Dog walking		1.01%
	Other		21.21%
Category error	Person walking a bicycle detected as person	25.05%	1.82%
	Person walking a shopping cart or the like detected as bicycle		11.52%
	Other (bicycle → person)		3.84%
	Other (person → bicycle)		7.88%

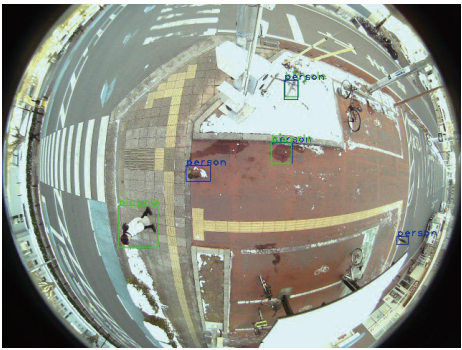
Although the initial model was subject to significant over-detection, one-cycle training of the current algorithm led to the almost total prevention of over-detection. Moreover, it should be possible to achieve performance improvements by training the model to be able to recognize respective separate categories regarding “infant with a parent,” “infant in baby stroller,” “motorcycle → bicycle,” “dog walking,” and “person walking a shopping cart or the like → bicycle” and by adjusting the threshold for object tracking regarding “person in bicycle parking area.”

Photo 4 presents an example of automatic annotation performed by the initial model. While the detection results produced by YOLO, shown in (a), determine a pedestrian as bicycle, the annotation results corrected the object to person to agree with Co-DETR. In addition, the annotation results also prevented various types of over-detection made by YOLO and Co-DETR. Note that the training data was generated properly although these images, captured on a day with remaining snow, represented a pattern not used in pre-training.

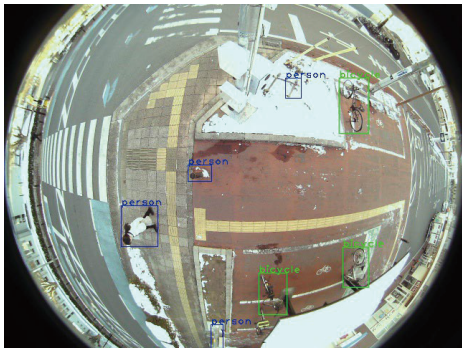
3-7 Summary of automatic performance improvements to pedestrian detection sensors

Automatic annotation was developed in order to improve the accuracy of image recognition AI. A demonstration test was conducted using images captured with fisheye lens at a public road intersection in Tsukuba City. The test results proved that it is possible to significantly improve the performance of trained models, with accuracy improving from 52.7% to 92.2% and detection rate from 55.7% to 94.2%.⁽¹⁾ Automatic annotation is particularly effective for the prevention of over-detection; and even one training cycle is satisfactorily effective. Furthermore, additional performance improvements are in sight.

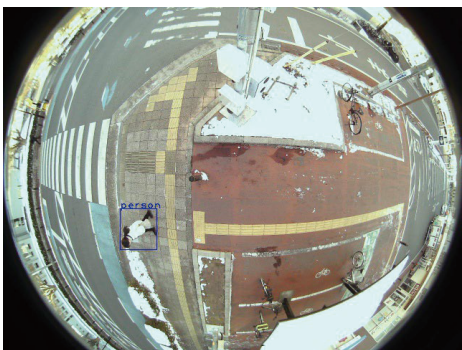
Note that in this demonstration the annotation results were directly used as training data; however, suitable images for training data can also be visually selected from



(a) Detection results produced by YOLO



(b) Detection results produced by Co-DETR



(c) Annotation results

Photo 4. Example of automatic annotation (using the initial model)

the annotation results. In that case, it is possible to significantly reduce work hours (to approximately two hours for 3,000 images).

4. Application for Factory Safety Improvements

In recent years, people have come to have increasingly higher factory safety awareness. Sumitomo Electric is engaging in various initiatives in order to live up to the “Safety First” slogan. In collaboration with SoftBank and TechnoPro, Inc. (hereinafter referred to as “TechnoPro”), it has conducted a demonstration experiment with the aim of improving the safety of forklift trucks and other industrial vehicles (hereinafter referred to as “industrial vehicles”) moving in factories.

In this demonstration experiment, a camera, a light emitting diode (LED) warning light, and the AI-equipped 5G terminal were set up, and camera images were sent to

SoftBank's multi-access edge computing (MEC) system via SoftBank's Private 5G (shared) channel. The MEC system had proximity detection AI created by TechnoPro running on it. The constructed system alerted everyone to danger by flashing the LED warning light in the factory when, upon image analysis, any approaching industrial vehicle was detected (A). Additionally, another system was constructed, which had the proximity detection AI created by TechnoPro transplanted to an AI-equipped 5G terminal, performed proximity detection by edge processing bypassing the 5G channel and MEC, and flashed a LED warning light (B).

Both systems were installed in the factory, and a demonstration experiment was conducted. As a result, it was proven that both configurations A and B detected approaching people and industrial vehicles and achieved the target latency time of 150 ms or less established to ensure a safe distance. Edge processing is anticipated to be applied in the safety field because it provides uninterrupted services even during a communications blackout.

5. Conclusion

An AI-equipped 5G terminal with AI capabilities was developed. In addition, demonstration experiments, conducted to test automatic performance improvements to pedestrian detection sensors and factory safety improvements, proved the validity of the AI-equipped 5G terminal.

6. Acknowledgements

This research report includes the achievements of the Cross-ministerial Strategic Innovation Promotion Program (SIP) Phase 3/Development of Smart Mobility Platform (research promoted by the New Energy and Industrial Technology Development Organization) (NEDO Reference No.: JPNP23023) promoted under the Council for Science, Technology and Innovation of the Cabinet Office.

Lastly, the authors are grateful to SoftBank Corporation and TechnoPro, Inc. for their cooperation in the demonstration experiment on factory safety improvements.



Technical Terms

- *1 Annotation: Labeling of image data with correct answers, and the labeled data, used to train AI.
- *2 Bounding box: The smallest rectangle that encloses a specific object in an image, which locates the object's position.
- *3 Confidence: Confidence indicates how accurately a bounding box encloses the object and the predicted probability for the category.
- *4 Precision: The evaluation index used in machine learning that indicates the proportion of truly correct answers relative to the positive (correct answer) predicted results produced by an AI model.

- *5 Recall: The evaluation index used in machine learning that indicates how accurate the prediction is relative to the truly positive (correct answer) results.

Reference

- (1) S. Maki, K. Yada, M. Furuta, S. Kadoi, T. Suzuka, and S. Ogawa, "Automatic performance improvement system for AI object detection models using images from a fisheye camera installed in a real environment," Vol.2026-CVIM-246 No.t1 (May 2026)

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

K. YADA*

• Senior Assistant Manager, Information Network Laboratory



S. MAKI

• Information Network Laboratory



M. FURUTA

• Information Network Laboratory



S. KADOI

• Information Network Laboratory



T. SUZUKA

• Assistant Manager, Information Network Laboratory



S. OGAWA

• Group Manager, Information Network Laboratory

