Fujitsu's Deep Learning Technology that Enables Smart City Monitoring

● Shuhei Kinoshita ● Masayuki Wakatsuki ● Shougo Horinouchi
● Satoshi Tanabe ● Akihiro Higashi

With low-cost IP cameras readily available across the world, the number of monitoring cameras installed in major cities has seen a dramatic rise in recent years. Smart City Monitoring based on AI is attracting attention as a means of making efficient use of these cameras. In March 2018, Fujitsu launched its new smart city monitoring solution FUJITSU Technical Computing Solution GREENAGES Citywide Surveillance V2. Citywide Surveillance is based on deep learning technology, and in order to apply this suite to a business operation, there are challenges to address in terms of accuracy, speed, and cost as fundamental features. Furthermore, as deep learning alone cannot deliver the value customers want, it is necessary to develop auxiliary technologies at the same time. This paper explains the Fujitsu technologies that bring customers the needed accuracy, speed, and cost performance as well as our co-creation initiatives to deliver value to our customers. It also introduces examples of Citywide Surveillance applied in business contexts.

1. Introduction

With low-cost IP cameras readily available, monitoring cameras have become prevalent all over cities. According to research by Yano Research Institute, 43.2 million monitoring cameras were shipped worldwide in 2018,[1] and in Japan the number is expected to reach 2.6 million by 2020 backed by the demand arising from upcoming international sporting events.[2] Against this background, there are growing needs to leverage monitoring cameras to ensure the safety and security of people, buildings and properties.

However, monitoring for security largely relies on human efforts at present; security operators must visually monitor the video images. This visual monitoring is difficult to make cost-efficient, and for this reason AI-enabled video monitoring (hereafter, Smart City Monitoring) has attracted attention in recent years. One of the anticipated features of Smart City Monitoring is the identification of suspicious or suspected individuals and vehicles by using all monitoring cameras installed in urban areas. Recently, use of Smart City Monitoring has extended beyond safety and security to marketing and road safety. For example, it can recognize frequent customers at a storefront and measure in-store traffic, congestion, and crowd-building automatically, which can be useful for route optimization and customer reception services.

It was in this context that Fujitsu launched its new Smart City Monitoring solution FUJITSU Technical Computing Solution GREENAGES Citywide Surveillance V2.

In this paper, we describe the technology for Citywide Surveillance and present examples of it used in business contexts.

2. Overview of Citywide Surveillance

In October 2016, Fujitsu launched Citywide Surveillance package V1 (version 1) as an intelligent monitoring suite based on deep learning and other machine learning technology applied to image/video analysis.[3] Subsequently, the system has been provided to both public and private sectors in Japan and abroad. With feedback and requests from many customers for additional features, we launched the second version (V2) in April 2018.

Citywide Surveillance V2 uses the following recognition and attribute classification functions based on video or photo images of individuals and vehicles from monitoring cameras (**Figure 1**):

1) Vehicles

Vehicles can be recognized in terms of their type, manufacturer, and body color. Other features include a vehicle registration number analysis, a vehicle counter for vehicles within a specified area or those passing predetermined lines or sections of lines set up on a screen.

2) People

People can be recognized in terms of the types and colors of their clothing and facial characteristics. Other features include a counter for people within a specified area or those who pass predetermined lines or sections of lines set up on a screen, as well as a feature to automatically estimate seating positions in restaurants, etc.

3. Challenges in applying deep learning to business

Citywide Surveillance is based on deep learning. Research on deep learning has gained impetus since the launch of Deep CNN[4] in 2012. In general, graphics processing units (GPU) are used for deep learning. But, there are still very few practical examples due to the fact that the required GPUs are very expensive and that it is still difficult to offer solutions to customers with deep learning alone. Therefore, we will focus below on the application of deep learning to business.

In our opinion, there are two major challenges in applying deep learning to business. The first challenge is to meet requirements in terms of analytical accuracy and speed, which directly concern the fundamental performance of deep learning, and of upfront installation costs. The second challenge is to develop and establish auxiliary technologies for coordinating deep learning technology into solutions.

1) Issues of accuracy, speed, and cost performance

Analytical accuracy is often considered an evaluation indicator for the results of deep learning operations. However, in a business context, the system's computing speed and GPU memory use are equally important factors. A high-performance GPU equipped with a large memory capacity is very expensive, hiking up the upfront costs, which customers must address.

2) Development of auxiliary technologies

Deep learning cannot provide solutions to customers by itself. Thus, developing deep-learning-assisted solutions will require auxiliary technologies to be developed and established. Candidate technologies for this purpose would be countless. Therefore, we have adopted an approach to co-create with customers such technologies that are directly relevant to their businesses.

4. Fujitsu technologies that address challenges of accuracy, speed, and cost performance

Fujitsu has various initiatives to adapt AI to businesses, leveraging the technologies it has mastered over the years. In this section, we describe Fujitsu technologies that address the challenges of accuracy, speed, and cost performance.

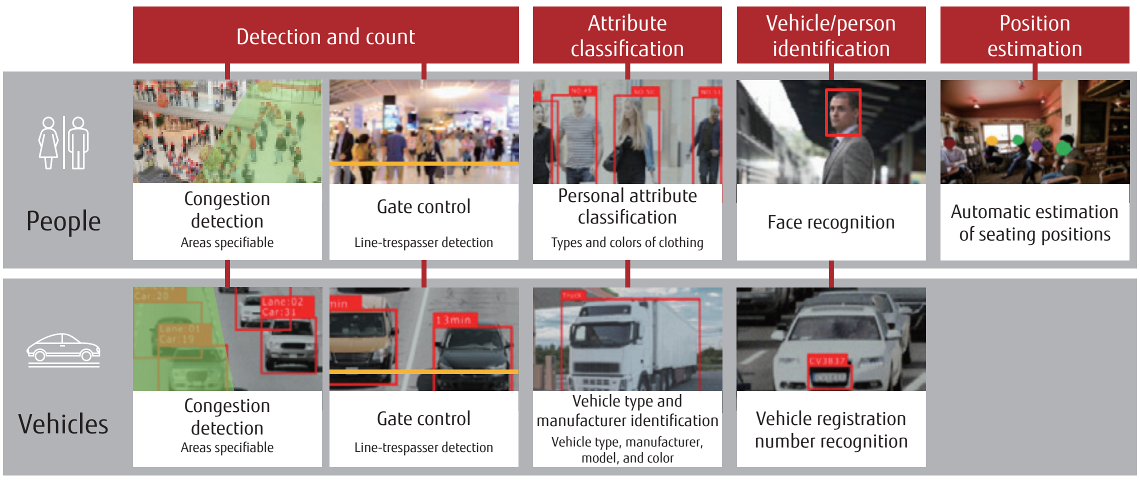

Figure 1
Analytical features of Citywide Surveillance.

4.1 Deep learning to identify colors with high accuracy

There are two approaches to enhance the accuracy of deep learning. The first is to improve learning models by, for example, multiplying the model layers. The second is to gather and learn from a lot of high-quality data. We take both of these approaches and improve learning models as well as enrich learning data. We describe in this section our attempt in the second approach, working on a technology to distinguish colors of vehicles through deep learning.

One of the difficulties in color discernment is that developers of learning data have individual difference in color perception. For example, one developer discerns the color of a vehicle in an image to be blue while another discerns it to be green. If learning data were prepared by different developers based on different standards, deep learning would be confused in the learning process, hindering improvements in accuracy.

We therefore organized color discernment rules based on the characteristics of the human eye, focusing on the three attributes of color,[5] namely, brightness, saturation, and hue. The impact of brightness, for example, is stronger for a color whose brightness is lower than a certain value. When saturation is lower than a certain value, hue becomes more difficult to perceive. Additionally, the tendency to be perceived as monotone becomes greater and the effect of saturation becomes stronger in such cases.

We developed and implemented the rules as shown in **Figure 2** as a tool for learning data developers, thereby succeeding in creating learning data that did not rely on individual judgment.

4.2 Speeding-up deep learning models by building small and efficient neural networks

To achieve both a reduction in GPU memory use and faster processing, it is important to make the deep learning model more efficient. In deep learning, most of the computation involves convolution. By reducing the number of convolution operations, it is possible to increase the processing speed. We first aggregated convolution layers of similar operations and created a common layer in order to reduce the number of convolution operations. This has halved the number of convolution operations needed to extract color properties.[6]

To further reduce the number of operations, we minimized the number of basic operations in the convolution operations. As convolution operations include matrix operations, the volume of operation can be reduced by approximating to the low rank matrices.[7] We initially secured a certain level of accuracy with the

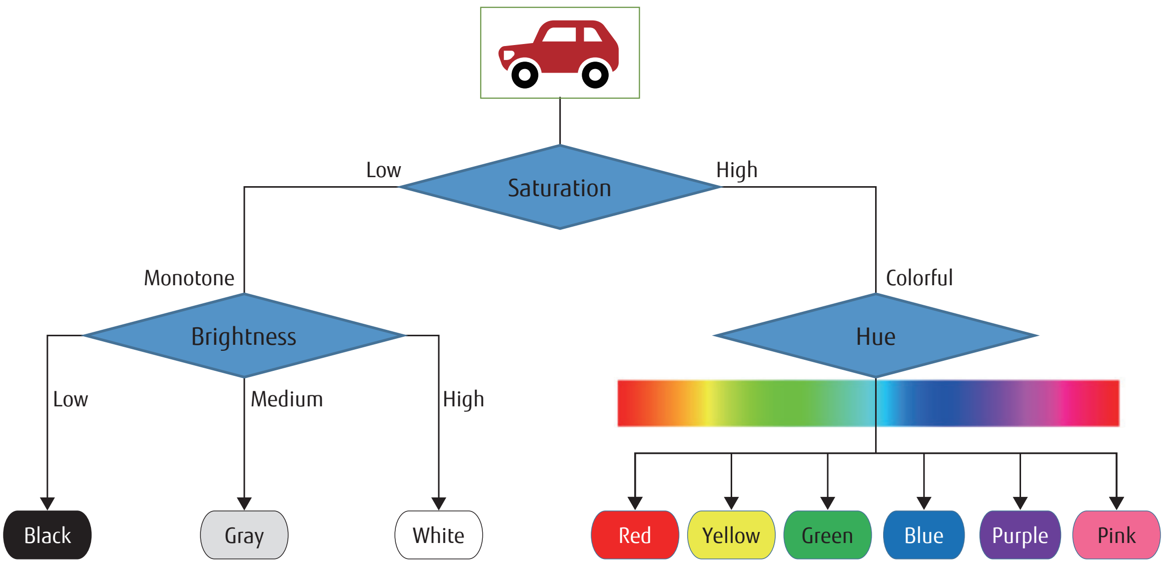

Figure 2
Color discernment rules.

convolution model (base model), then prepared another model (high speed model) using approximated low rank matrices. We then adopted stochastic gradient descent to achieve an overall optimization so that the intermediate outputs of these models are aligned. As a result of these efforts to make the models more efficient, we have succeeded in halving the process time without compromising the recognition accuracy.

4.3 Maintaining quality and preventing accuracy degradation of deep learning models

Having completed a learning model through various methods, efforts to further improve it continue, aiming to realize faster, more accurate models with a wider scope of analysis to meet customer needs. However, as deep learning processes are a black box, there is a risk of unexpected deterioration of accuracy. For example, feeding more learning data on the vehicles of a particular car manufacturer improves the accuracy in recognizing them, but it may compromise the accuracy in identifying similar vehicles of other manufacturers.

To solve this problem, it is necessary to re-validate the accuracy every time new learning data are added or the model is improved through evaluation. However, this process would be extremely costly if it were conducted through human effort. To address this problem, we have developed a tool that automates the process, from image analysis to the evaluation, aggregation, and visualization of the results. This tool has a feature that compares the deep learning results with the images of correct answers, automatically calculates the correct answer rate, and detects improvement/degradation against past performance. We also accelerated the deep learning model improvement by aggregating and analyzing the attributes in the correct answer data, such as the orientation of the vehicles.

Figure 3 illustrates the comparison of accuracy between Citywide Surveillance V2 and extant Fujitsu technology. A correct answer in this accuracy evaluation entails detection of a vehicle or a person and correct identification of various attributes. Citywide Surveillance V2 has significantly improved the accuracy on the extant technology. We will continue our efforts to improve the quality of the models and enrich the learning data to achieve even higher accuracy in future versions.

4.4 Adoption of cost-effective hardware

As a high-speed GPU equipped with a large memory capacity is very expensive, it is important to select appropriate hardware for customers according to their intended use of deep learning. Citywide Surveillance offers a selection of hardware according to different analysis targets or frequencies of use. The following are some of the features:

1) Users can easily turn off analytical features that they do not need to use and reduce the number of simultaneous analytical processes. This helps to reduce memory usage by the GPU.
2) By lowering the analysis frequency, input data from multiple cameras can be assigned to the same GPU and scheduled optimally.
3) For use that does not need real-time output, such as counting the number of people in unit areas or detecting congestion, the analysis can be processed using a CPU instead of a GPU.

By selecting the best hardware for the purpose required, it is possible to introduce the system with minimal investment on hardware.

5. Development of auxiliary technologies

The aforementioned deep learning technology alone is not sufficient to apply deep learning to business. It is often the case that the results returned by deep learning are not easy to comprehend. In order to achieve customer objectives, these results must be linked to customer value, for which the development

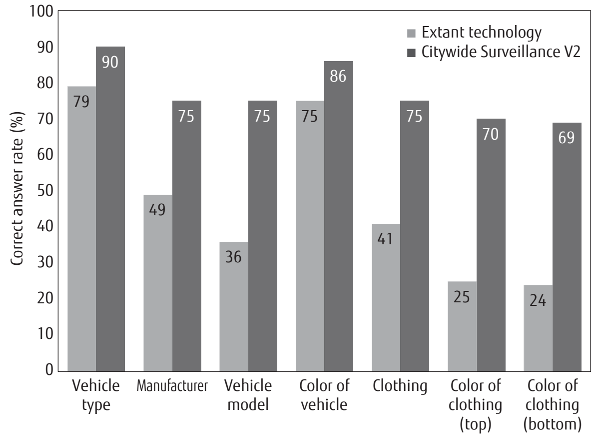

Figure 3
Evaluation of Citywide Surveillance accuracy.

of auxiliary technologies is indispensable. In intruder detection, for example, the system needs to have fast computing capabilities so that the detection of an intruder can be notified/indicated without delay.

5.1 Real-time scheduling

With ordinary camera images, it takes a few seconds from the time a vehicle or person enters the camera frame and disappears from it. For this reason, it is not necessary to perform the detection and identification operation for every single frame. In many cases, once every second is considered sufficient. In addition, performing the detection and identification for every frame would require significant amounts of computing, and it would be extremely difficult to follow the target in real time.

That is why Citywide Surveillance provides a real-time scheduling feature, which enables it to automatically track the vehicle/person detected in subsequent frames by not applying deep learning processes to all frames. Real-time analysis is divided into two processes: deep learning detection and identification, which needs a long time for computation, and tracking, which can be processed relatively quickly. We made real-time analysis possible by scheduling these processes at an interval optimal in terms of GPU performance, and reducing the overall processing time as a result.

5.2 Automatic estimation of seating positions

We collaborated with a major restaurant chain company to conduct a field trial of a system to manage table availability in a restaurant. At first, the system utilized deep learning to detect persons and compare the results against preset seating arrangement data to obtain the table occupancy rate. The client was happy with our system, which realized remarkably high accuracy compared to our competitors. On the other hand, they requested that the initial cost to introduce the system in many restaurants be reduced.

We thus developed a feature that automatically identified seating positions by detecting locations where persons remained unmoved for some time, as we noticed that persons remained in their seats for a certain period. The seating arrangement data obtained using the automatic seating position estimation technology and the visualization of available seats are shown in **Figure 4**.

Customers hesitate when making a decision to introduce a new system if it will create extra work for them. This automatic seating position estimation technology we developed enabled customers to obtain the seating arrangement data from multiple restaurants automatically without having to trouble their staff to prepare restaurant floor plan data. Furthermore, with this technology, it is now possible to respond flexibly to floor layout changes. In this way, we provided a customer with AI technology based on their needs and earned extremely high marks in terms of customer satisfaction.

Figure 4
Analysis results of table availability in a restaurant.

6. Use cases

Smart City Monitoring is expected to be leveraged in many aspects of urban life, making people's lives safer and more convenient, improving customer experiences, helping businesses to pursue management innovations, and so on. In this section, we present several Japanese and international use cases of operating Citywide Surveillance.

1) Japanese public institution

A system to efficiently detect criminal vehicles has been introduced to bolster security in parking areas and enhance the efficiency of security surveillance. The system can identify vehicle types and manufacturers based on images taken by many outdoor cameras installed on the premises. There is a plan to add a feature to identify persons to the system.

2) Overseas energy companies

We began offering services to detect frequent clients and black-listed individuals visiting their stores and to analyze trends in busy hours in stores, designed to help with sales promotion and enhance customer services. One of the clients awarded the system in

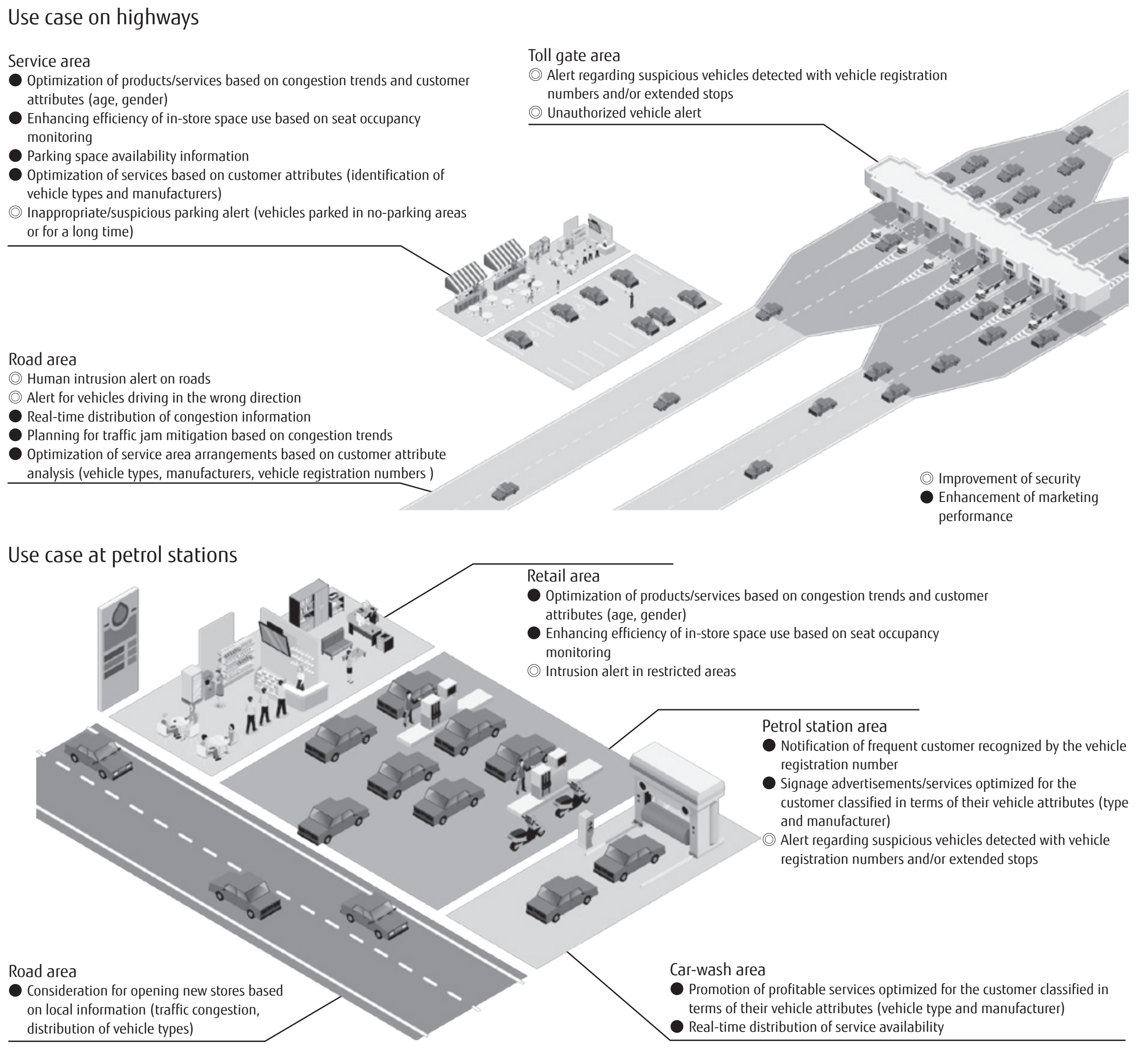

Figure 5
Use case of Citywide Surveillance V2.

2017 as the most innovative technology in recognition of the system's usefulness.

3) Japanese logistics company

The company is currently evaluating a system to count the number of entrances/exits for each load by trucks entering and exiting the distribution center. The purpose of this is to optimize vehicle circulation.

4) Japanese environmental surveillance organization

Aiming to automate traffic surveys, which conventionally rely on manual counting, the company is trying a service of automatic traffic surveillance (automatic count by vehicle type) using images taken by roadside cameras.

5) Japanese transport operator

Waiting in line to board or check-in on public transport is a major factor that dampens customer satisfaction. We conducted a trial to identify causes of congestion by monitoring crowded situations in real time using monitoring camera images.

6) Overseas transport operator

Visualization of congestion is important for public transport operators to improve customer satisfaction. We are currently conducting a field trial to analyze traffic congestion in urban areas based on images sent from cameras installed on route buses in operation.

There are many more situations in which Citywide Surveillance can potentially be leveraged in business beyond the above-mentioned cases. For example, the system can be used in combination with monitoring cameras installed on highways and/or petrol stations to identify traffic congestion. It also can contribute to the bolstering of marketing efforts such as optimized advertisements and the enhancing of security through the detection of illegally parked vehicles and vehicles driving in the wrong direction (**Figure 5**).

7. Conclusion

This paper described various efforts to use Citywide Surveillance in business based on deep learning.

Fujitsu will continue focusing on deep learning and other machine learning technologies to offer solutions that integrate vertically, from infrastructure to services. Through these efforts, we will work to solve a variety of social issues in the fields of urban infrastructure and security as well as industry and logistics.

References

1) Yano Research Institute: Surveillance Camera Market Forecast 2015. (in Japanese). *http://www.yano.co.jp/press/pdf/1420.pdf*
2) Yano Research Institute: Research on the Domestic IP Camera Market (2016). (in Japanese). *http://www.yano.co.jp/press/press.php/001599*
3) Fujitsu: Fujitsu Begins Sales of City Monitoring and Parking Management Solutions that Employ AI Technology. *http://www.fujitsu.com/global/about/resources/news/press-releases/2016/1003-01.html*
4) A. Krizhevsky, et al.: Imagenet classification with deep convolutional neural networks. Adv. In Neural Information Processing Systems (2012).
5) Wikipedia: Color appearance parameters. *https://en.wikipedia.org/wiki/Color_appearance_model#Color_appearance_parameters*
6) S. Tanabe, et al.: Real-Time Human Pose Estimation via Cascaded Neural Networks Embedded with Multi-task Learning. CAIP (2017).
7) M. Jaderberg, et al.: Speeding up Convolutional Neural Networks with Low Rank Expansions. arXiv preprint arXiv:1405.3866, 2015.

Shuhei Kinoshita
Fujitsu Ltd.
Mr. Kinoshita is currently engaged in the development of the GREENAGES Citywide Surveillance package.

Masayuki Wakatsuki
Fujitsu Ltd.
Mr. Wakatsuki is currently engaged in the development of the GREENAGES Citywide Surveillance package.

Shogo Horinouchi
Fujitsu Ltd.
Mr. Horinouchi is currently engaged in the development of the GREENAGES Citywide Surveillance package.

Satoshi Tanabe
Fujitsu Ltd.
Mr. Tanabe is currently engaged in the research and development of image analysis technology based on machine learning.

Akihiro Higashi
Fujitsu R&D Center Co., Ltd.
Mr. Higashi is currently engaged in the research and development of image analysis technology based on machine learning.