Application of Lightweight Deep Learning Technology to AISION®

Shozo Sakane Takamitsu Watanabe

"Image IoT-GW" is an edge device equipped with an image sensing function and lies at the heart of the image IoT system "AISION®*1)*2)." Application of lightweight deep learning technology to this image IoT-GW enables edge devices to perform high-precision image sensing. This article introduces an application example of lightweight deep learning technology to image IoT-GW that takes advantage of edge device features. In the description that follows, when describing general properties of an edge device, the device will simply be referred to as an "edge device," whereas during description of AISION features, the device will be referred to as "image IoT-GW."

Image Sensing using Deep Learning

IP network cameras have gained widespread use worldwide in recent years, and they are being used for various purposes. Among its uses, image sensing is attracting particular attention. **Table 1** shows the primary uses of image sensing.

Table 1. Primary Uses of Image Sensing

Classification	Recognition Objective	Use
1. Security Enhancement	Face	Detect specific person
	Behavior	Detect specific behavior, spot shoplifting
2. Marketing	Person	Count number of people in a specific area or passing through an area
	Personal Attribute	Extract personal characteristic and convert into attribute data such as age and gender
3. Work Efficiency	Object	Detect specific object, find misplaced object
	Vehicle Number	Extract license plate numbers and convert to data

Image sensing has been conventionally used with surveillance cameras to enhance security. However, in recent years, its application has spread to marketing and improving work efficiency, and its demand for use in these applications is expected to grow.

One of the reasons behind the spread of image sensing is the advancement in image analysis technology. In particular, the emergence of deep learning has dramatically improved the accuracy making image sensing more practical.

Comparison between conventional analysis technology and deep learning is shown in **Table 2**.

Table 2. Comparison of Conventional Analysis Technology and Deep Learning

Compared Topics	Conventional Technology	Deep Learning
Training of Analysis Methods	Humans (experts) teach computers knowledge about training data	Computer themselves learn from training data
Amount of Training Data	Small	Large
Accuracy with Respect to Amount of Training Data	Little improvement with additional training data	Big improvement with additional training data
Amount of Computation	Small	Large (especially with training)

Deep learning requires a large amount of training data to self-learn analysis methods and improve the accuracy of analysis, thus the amount of required computation is large leading to enormous computational cost.

Overview of Image Sensing with Image IoT-GW

The image IoT-GW provides various image sensing functions in a form of modules, and by selecting and installing the appropriate "image sensing module," it is possible to perform image sensing that matches the application. **Figure 1** shows an operational example of a image IoT-GW equipped with an image sensing module that measures traffic volume by vehicle type for designing and maintaining roads.

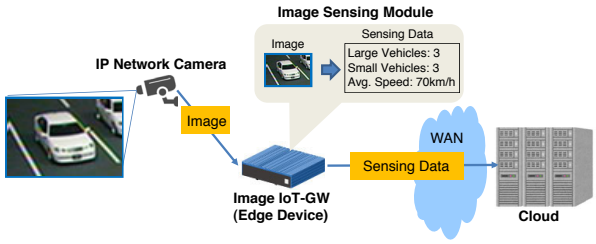

Figure 1. Image Sensing with Image IoT-GW (Vehicle Traffic Measurement)

Table 3 compares the features of performing image sensing in the edge device against image sensing in the cloud. In **Table 3**, (A) to (C) are advantages of the edge device and (D) to (E) are disadvantages. These are explained in detail below.

*1) AISION is a registered trademark of Oki Electric Industry Co., Ltd. *2) The name AISION is derived from AI (artificial intelligence, eye) and vision (insight, foresight, future)

Table 3. Comparison of Image Sensing between Edge Device and Cloud

Compared Topics	Edge Device	Cloud
(A) Processing Load	Distributed	Centralized
(B) Data Traffic	Small	Large
(C) Privacy Protection	Easy	Difficult
(D) Inter-image Interworking	Difficult	Easy
(E) High Load Analysis	Difficult	Easy

(1) Advantages of image sensing in edge device

In **Figure 1**, image of the vehicle traveling on the road is sent from the IP network camera to the image IoT-GW. The image IoT-GW performs image sensing on the received image, detects the vehicle, determines the type (large/small) and speed of the detected vehicle, and converts it into data. Performing image sensing in the image IoT-GW reduces the load on the cloud (center device) (**Table 3 (A)**). Moreover, by converting only the number of vehicles per type and the speeds to data at the image IoT-GW, it will be possible to send just a small amount of sensing data to the tabulation device (cloud) (**Table 3 (B)**). Image of a vehicle may contain personal information such as the driver's face and license plate, but after image sensing is performed, the image IoT-GW discards the image to prevent leakage of personal information (**Table 3 (C)**).

(2) Disadvantages of image sensing in edge device

Detecting/tracking people and vehicles (license plate) is an example of inter-image interworking (**Table 3 (D)**). Although people and vehicles can be detected at the edge device, in order to track where they are coming from and going, information from several locations is necessary. Therefore, the cloud is much better suited to handle the process. In case of high-load analysis (**Table 3 (E)**) such as facial recognition that requires identifying a specific person from a large amount of registered data, performing the process in cloud, which has a higher processing capability, is more efficient.

Overview of Lightweight Deep Learning

In deep learning, there is a "training" and "inference" process (**Figure 2**). "Training" optimizes the parameters of the neural network (such as the convolution filter and the connection weight of the fully connected layer) from the combination of large amount of images and correct answer data. The combination of the parameters obtained from training and the neural network structure is called a "trained model."

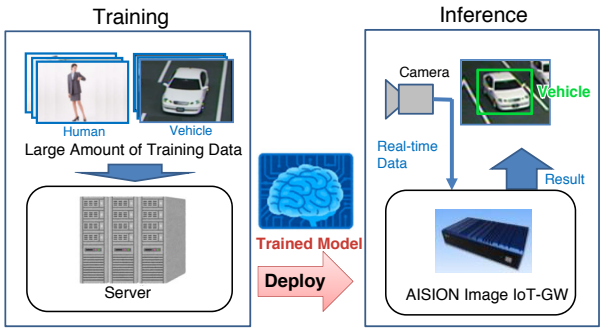

Figure 2. Training and Edge Device Inference

In "inference," input data is processed using the trained model. The inference computational cost and the amount of memory depend on the depth of the model layer and number of parameters. Therefore, when inference processing is performed in the edge device, a lightweight model is required to operate with limited calculation resources.

When deep learning inference process is performed in the edge device, processor capacity and memory amount become problems. In the image IoT-GW, the effort to produce a lightweight model has made it possible to install an image sensing module that utilizes deep learning. Below is a description of lightweight deep learning implemented in the image sensing module under development.

(1) Reducing model parameters

Computational cost and memory amount can be reduced by eliminating redundant parameters that do not affect inference accuracy from the trained model. As shown in **Figure 3**, deep learning for image recognition consists of a convolution layer or pooling layer and a fully connected layer. In the convolution layer, several filters are applied to an image to extract features, but by reducing redundant filters, computational cost can be mainly suppressed. Additionally, in the fully connected layer, classification is made based on the amount of features extracted in the convolution layer, but by reducing parameters with low-rank approximation[1)] to reduce the weight of the fully connected layer, memory amount can be mainly suppressed. Here, low-rank approximation is a weight reduction method of compressing the dimensions of a matrix using singular value decomposition or the like.

In the image sensing module under development, the computational cost and the memory amount are reduced by reducing the parameters of the convolution layer and the fully connected layer to the extent that the accuracy does not deteriorate.

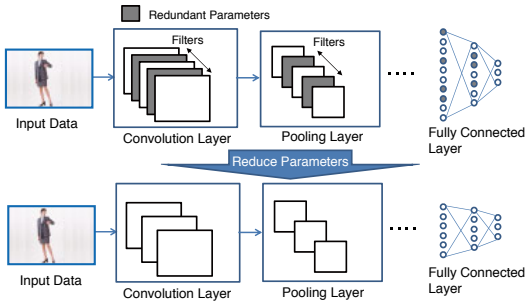

Figure 3. Reduction of Model Parameters

(2) Low-bit quantization

Deep learning computation generally performed with 32-bit precision can be replaced with low-bit computation of 16 bits or less to reduce memory usage and speed up computation regardless of model structure. However, there is a trade-off between the number of bits used in the computation and inference performance, and performance deterioration will occur in a low-bit quantized model compared with the original model. Performance degradation is especially noticeable with regression problems such as estimating age from facial images. Therefore, it is necessary to determine the appropriate bit quantization depending on the object to be inferred with deep learning.

(3) Knowledge distillation

In deep learning, the inference accuracy tends to improve as models' layers deepen and parameters increase. On the other hand, lightweight models with shallow layers and few parameters are suitable for edge device operation. Hence, as shown in **Figure 4**, considering a model with high inference accuracy but large computational cost as a "teacher" and a lightweight model operable in an edge device as a "student," the class probabilities of the teacher's middle and output layers are taught to the student in a method called "knowledge distillation[2]."

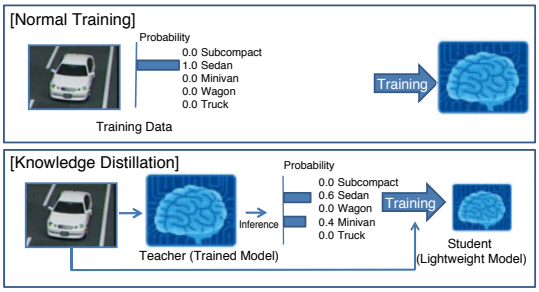

Figure 4. Knowledge Distillation

For example, a vehicle type that is hard to classify as a "sedan" or "minivan" even by humans should not be taught to be a "sedan" as in a normal training. Instead, teaching the student with the teacher's output of "sedan resembling a minivan" will enable a more consistent training. Conducting knowledge distillation on the image sensing module under development has resulted in the same accuracy as normal training, but at 1/3 the computational cost.

(4) Frequency superimposition

The techniques described so far are common methods of lightweight deep learning adopted in the image IoT-GW's image sensing module. Finally, OKI's original frequency superimposition technology will be introduced. OKI has developed "frequency superimposition deep learning," which superimposes the spatial frequency analysis result and luminance information of the input image. This enables the lightweight model to be tolerant against adverse environment (**Figure 5**). Here, the spatial frequency represents how often the intensity of an image changes. Objects off the focal point of the camera appear blurred and many low frequency components are present. On the other hand, focused objects appear sharp and high frequency components increase. Snow and rain that cross the camera lens during bad weather are often blurred, therefore under such conditions, removing low frequency components improves the identification accuracy. Moreover, superimposing the high frequency components lost due to image reduction enables high identification accuracy to be maintained even if input data size is reduced. In vehicle identification evaluation using actual images under various weather conditions, it was confirmed that processing speed can be increased 16 times and memory amount can be reduced to 1/10 while increasing identification accuracy from 98.0% to 99.8%.

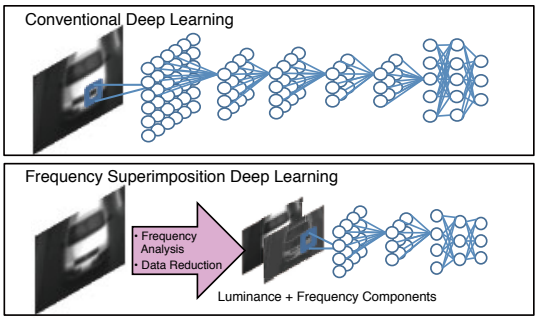

Figure 5. Frequency Superimposition

Issues and Future of Deep Learning with AISION Image IoT-GW

Up to this point, lightweight deep learning has been described. Although an edge device is suited for real-

time inference processing due to its closeness to the sensors, it is not suitable for training that requires huge computations (**Table 4**).

However, in order to improve deep learning to the accuracy suitable for a site, it is desirable to train with field data while under actual operation or PoC[*3].

Additionally, for applications where data is frequently added such as personal identification through facial images and model identification of vehicles, it is necessary to update trained models as needed with respect to edge devices making management complicated.

In response to these issues, OKI is studying to interwork the image IoT-GW with the cloud.

Table 4. Characteristic of Edge Device and cloud

	Edge Device	Cloud
Distance to Sensors	Near	Far
Computational Capacity (standalone)	Low	High
System Management	Distributed	Centralized
Data Traffic	Small	Large

(1) Adaptation to field environment through cloud interworking

Inference can be adapted to a field environment by collecting field data with the image IoT-GW and training in the cloud.

Figure 6 shows an example of adapting deep learning to a field environment in which multiple image IoT-GWs and the cloud are interworked. Field data is sent from the image IoT-GW to the cloud, and in the cloud, models are retrained using field data collected from multiple image IoT-GWs. The trained models are redeployed to the image IoT-GWs. This cycle is continuously repeated to evolve the models so that their adaptation as trained models for the particular field environment will improve.

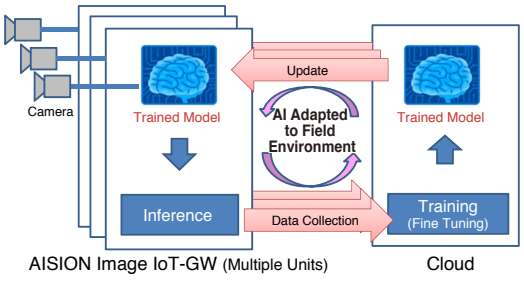

Figure 6. Adaptation to Field Environment

(2) Distributed inference processing through cloud interworking

It becomes possible to improve inference accuracy and maintainability when the task of inference processing is distributed between the cloud and image IoT-GW.

Figure 7 shows an example of distributed processing in which the image IoT-GW interworks with the cloud. The image IoT-GW performs primary inference processing (example: face detection), and in the cloud, secondary inference processing (example: personal identification) is performed with a deep layered model. If the entire processing is performed in the cloud, it is necessary to send image data to the cloud, which would increase the traffic load on the communication network. However, distributing the processing between the image IoT-GW and the cloud, the amount of communication can be reduced and higher inference accuracy can be achieved as opposed to standalone processing in the edge device. Furthermore, maintenance of the image IoT-GW is also reduced by performing inference that requires frequent retraining in the cloud.

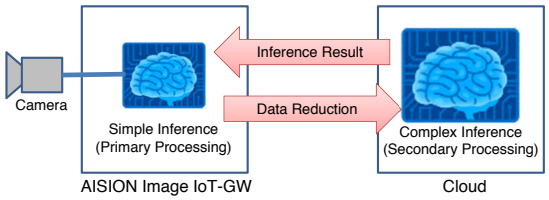

Figure 7. Distributed Processing with Cloud

Conclusion

This article introduced lightweight deep learning technology and its application to the image IoT-GW. There are no plans to replace the entire image sensing provided in the image IoT-GW with deep learning. It is important to consider the advantage/disadvantage of deep learning and apply deep learning to the necessary parts based on the requirements of the application. In order to expand the use of image sensing, OKI plans to provide a high-precision image sensing module utilizing advanced analytical technologies such as lightweight deep learning. ◆◆

References

1) J. Xue, J. Li, and Y. Gong, "Restructuring of Deep Neural Network Acoustic Models with Singular Value Decomposition," INTERSPEECH, pp. 2365-2369 (2013)
2) Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the Knowledge in a Neural Network," arXiv preprint

Authors

Shozo Sakane, Systems Development Department-5, Network Systems Division, ICT Business Group

Takamitsu Watanabe, Advanced Technology Development Department, Fundamental Technology Center, ICT Business Group

*3) PoC (Proof of Concept): Simple trial to verify the feasibility of new concepts and theories.