Technology Reports (Special Articles) AI Image Recognition Deep Learning

Special Articles on AI—Expansion of AI Technologies to Diverse Industries and Basic Technologies Supporting AI Applications—

Providing Image Recognition AI via the DOCOMO Image Recognition Platform

Service Innovation Department Toshiki Sakai Motoki Iwata

In recent years, AI technology has become widespread, and in the field of image recognition, it is being used to replace, automate, and save human labor. However, its introduction has necessitated the preparation of high-performance server environments, the installation of various software packages, and the creation of AI development environments. For this reason, we have developed the DOCOMO Image Recognition Platform to facilitate the development and deployment of diverse image recognition AI systems. This has made it possible for AI users to introduce image recognition AI technology simply by preparing data for image recognition AI instead of having to prepare their own AI development and operation environments.

1. Introduction

In recent years, AI technology based on deep learning*1 has made remarkable advances and has become very popular. Particularly strong progress has been made in industrial applications of image recognition AI*2. For example, it is being used to support or even replace tasks that have hitherto been done by humans, such as assisting with product inspections in factories, detecting people and vehicles in security camera images, or supporting diagnostic imaging in medical care. However, when image recognition AI is used for the streamlining, automation and Digital Transformation (DX)*3 of

*1 Deep learning: A machine learning method that can learn more complex concepts and perform judgments and estimations by using a more complex form of neural network, which is designed to imitate human neural processing mechanisms.

*2 Image recognition AI: A form of AI that takes images as its input and uses them to generate results by making judgments, performing estimates, and so on.

existing jobs, it is seldom available in a form that can directly solve the issues faced by each company. For this reason, each company must first prepare its own collection of images and/or video data together with annotations that record the answers corresponding to the type of image recognition processing the AI is needed to perform on the data, and must then perform training to develop its own image recognition AI, followed by deployment*4 in order to make the trained image recognition AI usable. Deployment involves setting up this image recognition AI on a server and building an Application Programming Interface (API)*5 that inputs images and outputs results as text or the like. However, for AI users to perform this training and deployment by themselves, there are several hurdles they have to tackle, including understanding the framework*6 used for working with deep learning, performing high-speed training, and preparing hardware for recognition.

To make things easier, services are available that facilitate the training and deployment of image recognition on the cloud. Examples include Amazon Rekognition Custom Labels [1], and AutoML Vision [2]. In 2020, we released the DOCOMO Image Recognition Platform [3] to make it easier for more users to access the image recognition*7 technology we developed in-house.

When providing functions for training and deploying image recognition AI, since each user requires different image recognition functions, a mechanism is required for providing these functions more efficiently. This article describes the mechanism used to provide multiple image recognition functions on the DOCOMO Image Recognition Platform. The DOCOMO Image Recognition Platform also provides a more secure image recognition environment by performing image recognition within the DOCOMO's closed network, as will be explained below.

2. Overview of the DOCOMO Image Recognition Platform

The DOCOMO Image Recognition Platform provides functions for training image recognition AI in the cloud and deploying it in a state where inference using image recognition AI can be performed.

1) Image Recognition Functions

Figure 1 shows the image recognition functions provided by the DOCOMO Image Recognition Platform: (1) object detection (find a specific object in an image and estimate its coordinates within the image), (2) generic object recognition (classify objects into categories based on features of the objects themselves, their surroundings, the scene, and the image as a whole), (3) character recognition (recognize text characters in an image), (4) similar image search (search for images similar to a given image), (5) specific object recognition (identify a specific object by matching an image of this object with multiple pre-prepared images of objects), and (6) pose estimation (estimate the position of a person's skeleton and joints in an image).

2) Assumed Use Cases

Figure 2 shows the assumed use cases of each image recognition function. Object detection can be used to detect weed growth in drone images and people and cars in surveillance camera images.

*3 DX: The changes that digital technology causes or influences in all aspects of human life.

*4 Deployment: Installing applications by placing them in their execution environments.

*5 API: An interface that enables the functions of software to be used by other programs.

*6 Framework: Software that encompasses functionality and control structures generally required for software in a given domain. In contrast to a library in which the developer calls individual functions, code in the framework handles overall control and calls individual functions added by the developer.

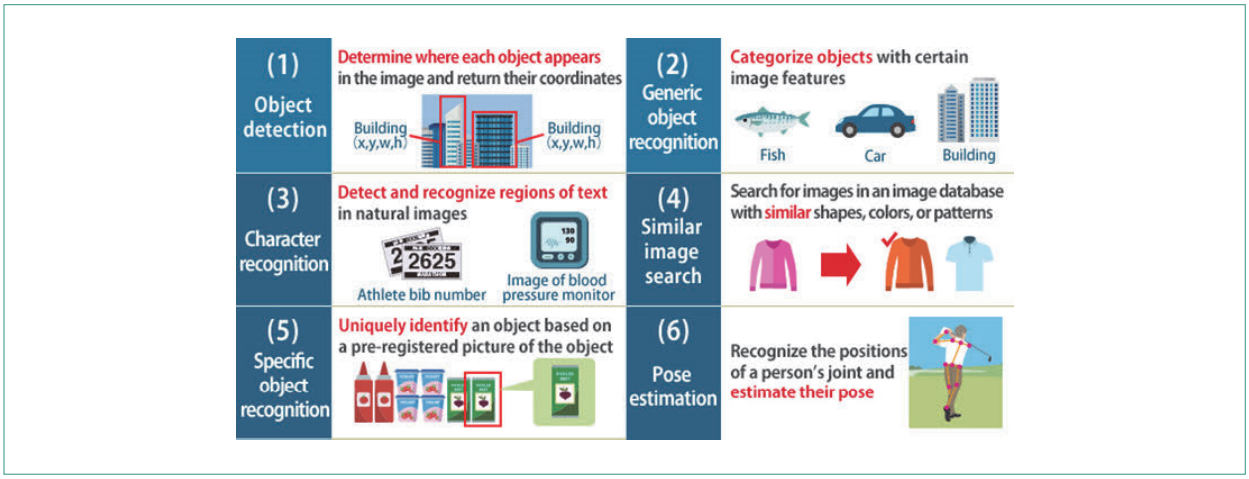

Figure 1 Image recognition functions provided by the DOCOMO Image Recognition Platform

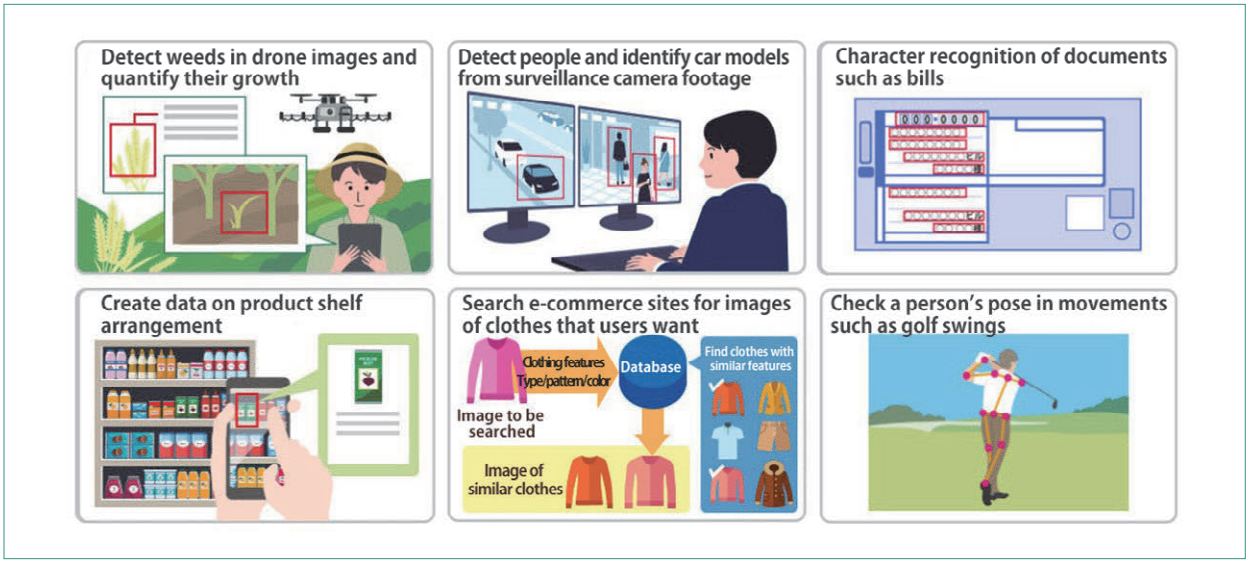

Figure 2 Assumed use cases of each image recognition function

By combining generic object recognition with the results of detecting areas that contain plants or cars, it is also possible to determine the amount of plant growth and the types of cars. Another possible use is to combine object detection and character recognition to read characters from bills and documents, or to combine object detection and specific object recognition so that images of product shelves can be converted into data on what products are stored where.

In addition, similar image search can be applied to images of fashion items in order to find similar items, and pose estimation can be used to check the form of actions performed by sports players.

*7 Image recognition: Technology that uses image processing and machine learning (see *8) to enable machines to understand images and extract meaning from them.

3) Functional Configuration

The configuration of the DOCOMO Image Recognition Platform is shown in **Figure 3**. For object detection and generic object recognition (category classification), the platform provides both training and inference functions as custom training models that can implement tailor-made AI based on data provided by the user. In addition, for similar image search and specific object recognition, it provides a user dictionary creation function that enables searching and recognition based on the user's own image data. For object detection, generic object recognition (category classification), character recognition, and pose estimation, NTT DOCOMO also provides pre-trained image recognition AI (common pre-trained model) that was trained by NTT DOCOMO so that users don't have to prepare their own training data to make inferences.

4) Provision of Website/web Console

The DOCOMO Image Recognition Platform provides users with a website/web console. By accessing this console, users can train their AI, check the results of this training (evaluation), and deploy the image recognition AI that they have created and trained. Users can input recognition requests (inference requests) to the deployed AI through a WebAPI interface. When image data is input to the WebAPI, the recognition results are returned in text format.

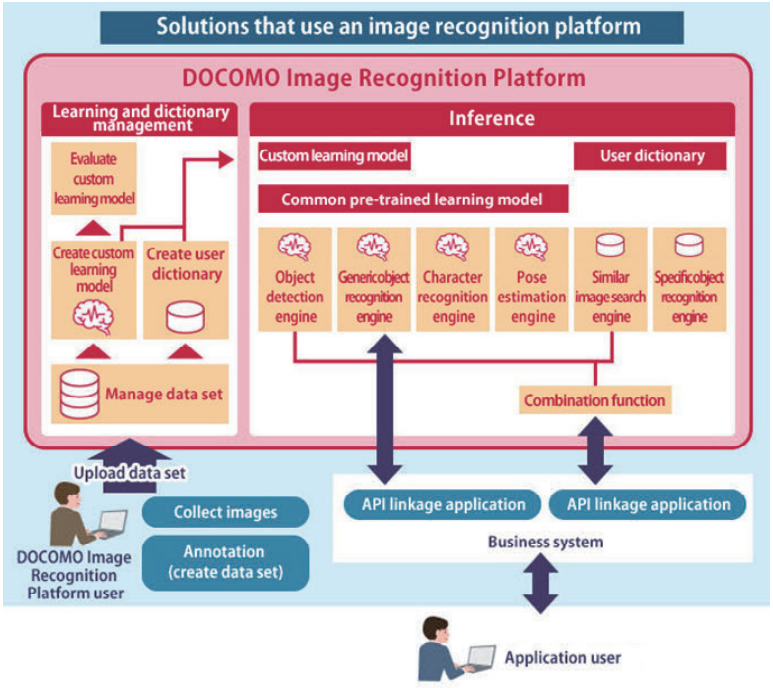

Figure 3 Functioning of the DOCOMO Image Recognition Platform

3. Two Initiatives on the DOCOMO Image Recognition Platform

We are working to make the DOCOMO Image Recognition Platform more convenient for users through two initiatives. The first is to containerize the image recognition functions so that they can be provided, developed and updated more quickly. This makes it possible to provide a full lineup of image recognition functions, enabling the provision of functions that can solve users' problems. The second is to build the DOCOMO Image Recognition Platform on a data center within the DOCOMO's closed network, so that it can be accessed without sending anything over the Internet. This makes it possible to provide secure image recognition AI.

3.1 WebAPI for Image Recognition AI Using Containers

As mentioned above, the DOCOMO Image Recognition Platform provides multiple image recognition functions. The image recognition functions required by users vary depending on the problem they want to solve and are expected to become more varied in the future. Advances in image recognition AI technology are being made every day, and to provide users with better accuracy and higher speeds, the image recognition functions that have already been deployed must be regularly updated. Therefore, to speed up the provision, development and updating of each image recognition function in the DOCOMO Image Recognition Platform, the functions are virtualized using containers and the containerized functions are modularized and shared, as illustrated in **Figure 4**.

Container-based virtualization is a technology whereby applications such as image recognition functions are combined with the libraries needed to run them in the form of packages called containers, allowing these applications to run with minimal

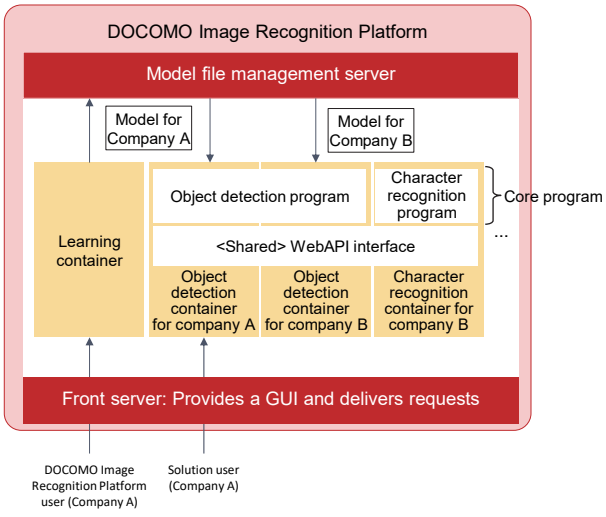

Figure 4 Using containerization to provide image recognition functions

dependence on the server's OS/environment. In the DOCOMO Image Recognition Platform, each image recognition function's core program and its WebAPI interface are combined in a single container. In this way, it is possible to add new image recognition functions, update existing image recognition functions, and augment the recognition resources (changing them so that taining and inference can be performed from more images) by adding/removing containers within the DOCOMO Image Recognition Platform.

Furthermore, within an image recognition container, the program that handles the abovementioned interface is separate from the core program for image recognition, and by using a shared program with a standardized interface, less work needs to be done to develop new image recognition functions. The input/output format of the interface between the interface program and the core program can be flexibly changed from within each core program as shown in **Table 1** to ensure it has the flexibility to enable the expansion of image recognition functions.

In addition, the image recognition models trained for each image recognition AI user are stored independently outside the container. This makes it possible to augment only the resources of a specific user-generated image recognition AI, or to replace just a single container and inherit its model when a container is updated.

3.2 Secure Image Recognition Using the DOCOMO Open Innovation Cloud

The use of image recognition may require the input of sensitive image data. To prevent images from being released to the outside world when performing image recognition, it is important to consider the security of the transmission paths used to transfer these images, and of the servers where the image recognition processing is performed.

When performing image recognition, the image recognition system can be implemented in the cloud or in its own data center to consolidate the processing functions. This makes it possible to increase the utilization of computational resources in the cloud or data center so that image recognition processing can be performed by making efficient use of these resources. When doing so, the risk can be reduced by equipping the system with appropriate security measures. However, even in this case,

Table 1 Division of roles between the image recognition interface program and the core program

	Standardization	Defined in each core program
When the container starts	· Trained image recognition model · How to input a file	· How to load a trained image recognition model
Input during inference	WebAPI input format · JSON format · Multipart/form-data format	WebAPI input details · How to store images in WebAPI input requests, etc.
Output during inference	WebAPI output format · JSON format	WebAPI output details · How to store image recognition results using JSON format in WebAPI output

the route over which images are sent to the cloud or data center needs to be separately secured, and steps must be taken to secure this route by without going via the Internet, such as by using a leased line.

One way to address this issue is to perform the image recognition process on a local PC, smartphone, or edge computing device. This prevents images from being transmitted over the network and keeps the images secure because they stay within the device. On the other hand, edge computing devices, PCs and smartphones are insufficiently powerful to perform image recognition, so the recognition process takes a long time. Furthermore, when image recognition is performed on individual devices, even if they do have sufficient performance, it is not possible to consolidate these image recognition processes to increase the efficiency of resource utilization compared to processing performed in the cloud or in a data center.

The DOCOMO Image Recognition Platform solves these issues by setting up a data center (the DOCOMO Open Innovation Cloud) within the NTT DOCOMO communication network as an intermediate between the two, where image recognition processing can be performed. A conceptual illustration of this configuration is shown in **Figure 5**. Since the image recognition process is done in a data center within the NTT DOCOMO communication network, it is possible to communicate with this data center without sending anything via the Internet provided it is accessed via a NTT DOCOMO 4G or 5G line (4th or 5th Generation mobile communication system). This can reduce the risk of image information being leaked to the outside world. On the other hand, since the image recognition functions are centralized at the data center, it is also possible to optimize the computational cost.

In the DOCOMO Image Recognition Platform, connections using this closed network are provided as a Cloud Direct [4] connection option. An additional advantage of this closed network access is that it bypasses the Internet, which means that connections can be made with a shorter delay. The

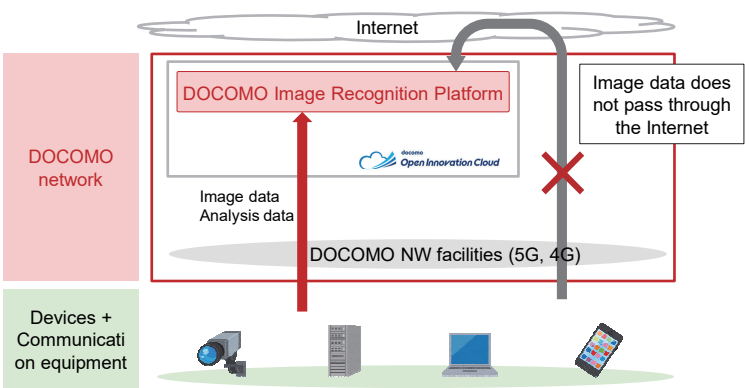

Figure 5 Using closed network connections to make image recognition secure

use of 5G for these connections can reduce the delays still further.

4. Building and Using an Image Recognition API on the DOCOMO Image Recognition Platform

In the DOCOMO Image Recognition Platform, the steps that need to be performed before using the API can be broadly divided into training and deployment. With this service, it is possible to perform these tasks on a web browser (in the cloud) with a simple user interface (**Figure 6**). Training involves creating a model specific to each individual task, and deployment involves using either a trained model or a general-purpose model to perform image recognition. In addition, users can create their own dictionaries for similar image search and specific object recognition functions. A detailed manual for each process is available on the DOCOMO Open Innovation Cloud developer portal.

4.1 Implementing Training

In the current DOCOMO Image Recognition Platform, training can be performed in the object detection and generic object recognition functions. The specific training procedure is described in the tutorial in the developer portal.

In addition to image data, the training process also requires annotation data. This is data that has been tagged with information necessary for the training and evaluation processes of general machine learning*8, including this image recognition technique. In the training process, the model is

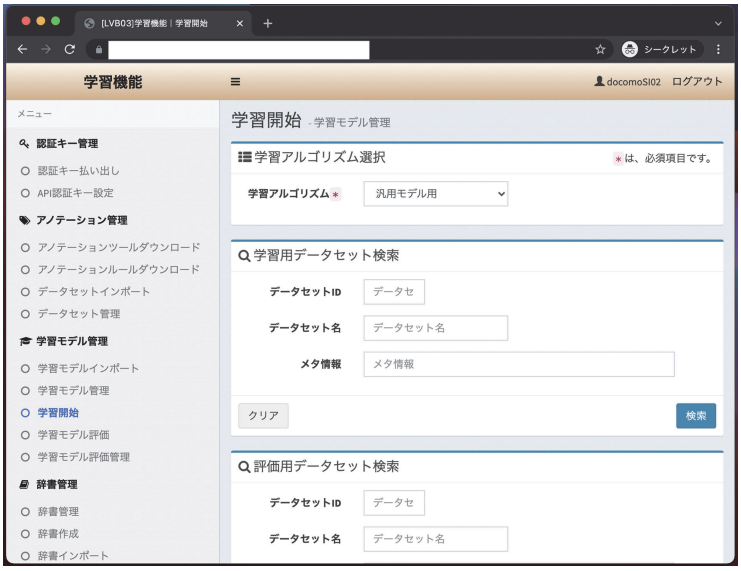

Figure 6 Service admin screen

*8 Machine learning: A technology that enables a computer to learn useful judgment standards through statistical processing from sample data.

trained using annotation data as examples of correct information, and in the evaluation process, the model is evaluated by comparing its inference results with the annotation data. A tool for creating annotation data is provided by NTT DOCOMO as a sample tool.

In the local environment, the user creates a data set consisting of image data and corresponding annotation data according to the specifications and uploads it to the DOCOMO Image Recognition Platform. A verification data set is also created and uploaded in a similar manner, and training is performed by setting up the training data set, verification data set, and target task (object detection or generic object recognition).

Trained models can be evaluated for accuracy on the platform. Like the data sets used for training and verification, an evaluation data set is created and uploaded, and then evaluation is performed by selecting a trained model and an evaluation data set from "Trained model evaluation" in the service admin screen. The evaluation results can be downloaded from the platform. From these results, it is possible to check the indexes showing the accuracy[*9], precision[*10], recall[*11], and F1-score[*12] for both image classification and object detection.

4.2 Implementing Deployment

Deployment makes it possible to use (i.e., perform inference using) models created by training and models independently trained and provided by NTT DOCOMO via a WebAPI. A deployed image recognition AI can be used by dispensing and assigning an API key. A REpresentational State Transfer (REST) API[*13] is adopted as the API method and can be linked with the user-side system with a simple design.

A deployed image recognition AI can be managed by adjusting the API management and API authentication key settings in the service admin screen. In API management, it is possible to manage trained models and add or delete image recognition AI. These functions can be easily scaled. In API authentication key setting, it is possible to assign and change the authentication key of a deployed image recognition AI.

4.3 Implementing Dictionary Creation

On the DOCOMO Image Recognition Platform, it is possible to create unique dictionaries for similar image search and specific object recognition. In similar image search, it is possible to determine which images in a pre-prepared dictionary of images resemble the requested image. Furthermore, specific object recognition can identify what is depicted in the requested image by comparing it with images in a pre-prepared dictionary.

The preparation of both types of dictionary involves collecting and annotating images. After a dictionary has been created, each function can be used by deploying it in the same way as other image recognition functions.

5. Conclusion

In this article, we described the background and challenges of making image recognition services easy to use, and we discussed the image recognition functions provided by the DOCOMO Image Recognition Platform. We described the character-

*9 Accuracy: The percentage of inferred data that is correctly classified and detected.
*10 Precision: In object detection, the percentage of detected objects that are certain to be objects corresponding to a given label. In generic object recognition, the percentage of data inferred to correspond to a certain label for which these inferred results are correct.
*11 Recall: The percentage of all data for a label that is correctly classified with that label (in object detection).
*12 F1-score: The harmonic mean of precision and recall.
*13 REST API: An API that adheres to the REST style of software architecture, which evolved from design principles proposed by Roy Fielding in 2000.

istics of the system that enable the provision of various image recognition functions in the DOCOMO Image Recognition Platform, and way in which it improves the security of image recognition by performing recognition processing in a closed network. We also showed how the DOCOMO Image Recognition Platform is actually used in practice. Going forward, NTT DOCOMO will continue to improve and update its functions to provide the image recognition functions that users need.

REFERENCES

[1] AWS: "Amazon Rekognition Custom Labels." https://aws.amazon.com/rekognition/custom-labels-features/?nc1=h_ls
[2] Google Cloud: "AutoML Vision." https://cloud.google.com/vision/automl/docs
[3] NTT DOCOMO: "The DOCOMO Image Recognition Platform." https://www.nttdocomo.co.jp/biz/service/dirp/
[4] NTT DOCOMO: "What is Cloud Direct?" https://developer.dev-portal.d-oic.com/document/docs/cloud-direct/concepts/overview.html