|  |
| --- |
|  |

|  |  |
| --- | --- |
| Input Contribution | |
| Meeting ID\* | RDM#55 |
| Title:\* | Key issue on data augmentation |
| Source:\* | JaeSeung Song, Sejong University, jssong@sejong.ac.kr |
| Date:\* | 2022-07-11 |
| Input related to\* | TR-0068 (Rel-5) |
| Intended purpose of  document:\* | Decision  Discussion  Information  Other <specify> |
| Impacted other TS/TR(s) |  |
| Decision requested or recommendation:\* | Agree for inclusion in TR-0068 |
| Template Version: January 2017 (Do not modify) | |

**oneM2M Notice**

The document to which this cover statement is attached is submitted to oneM2M. Participation in, or attendance at, any activity of oneM2M, constitutes acceptance of and agreement to be bound by terms of the Working Procedures and the Partnership Agreement, including the Intellectual Property Rights (IPR) Principles Governing oneM2M Work found in Annex 1 of the Partnership Agreement.

# Introduction

This contribution introduces a key issue on data augmentation.

R01:

* Split this part into two parts (first sub-section for key issue and the second sub-section for potential solution or high-level ~~~~ ) 🡺 Potential (preferable one) vs Possible (something might happen)

### -----------------------Start of change 1-------------------------------------------

## 8.2 Key Issue & Possible Concept 1 – Data augmentation

### 8.2.1 Key Issue

Many AI applications use data collected in IoT platforms. However, AI services often do not produce accurate results due to insufficient training data. Data augmentation in AI/ML, especially supervised machine learning, is a technique used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data.

For example, an AI application trains its model based on images stored in various resources in an IoT platform. The AI application can enhance the existing data set if there is a small dataset of images to build a better model. Typically, the AI application can take a small dataset of images and transform the objects to different sizes by zooming in or zooming out, flipping them vertically or horizontally or changing the brightness, whatever makes sense for the object.

Currently, the oneM2M platform does not provide any features supporting AI/ML applications to have increased data for AI/ML training. If the oneM2M platform support such data augmentation features, AI applications can easily build a model based on an augmented dataset.

### 8.2.2 Possible Solution

For convenience, this chapter considers that there exists a small dataset of images in the oneM2M platform for AI/ML training. Then the oneM2M platform provides features to support the following data augmentation to its image resources:

* Flipping: flipping the image vertically or horizontally
* Rotation: rotates the image by a specified degree
* Cropping: objects appear in different positions in different proportions in the image
* Zoom in, Zoom out
* Changing brightness, contrast, colors, etc

For each augmentation type, the AI application requires different configuration parameters. For example, the following information is required to perform data augmentation:

* Source resource URI that contains a target image
* Types of data augmentation (e.g., flipping, rotation, and cropping)
* Parameters for the selected data augmentation
* A destination resource URI to store a dataset of augmented images

For example, a data augmentation with the rotated image type can be performed by specifying a rotation degree and the total number of images to generate. Suppose there is a request from an AI/ML application to perform image rotation with input parameters, rotation 1 degree and generate 90 augmented images. In that case, the oneM2M platform generates 90 images by rotating the source image by 1 degree 90 times.



Figure 8.2-1: An example structure of [*dataAugmentation*] resource



Figure 8.2-2: High-level procedure to perform data augmentation

This can be done by introducing a new resource called the <dataAugmentation> to hold the information required to perform data augmentation. The <dataAugmentation> resource can have the following attributes (see Figure 8.2-1):

* Type: type of data augmentation (e.g., resize, crop, rotate)
* Source resource: a resource that contains the raw image
* Augmentation parameter: required parameters for the selected augmentation type
* Target resource: a resource or a set of resources to store generated images

Figure 8.2-2 shows a high-level procedure that an AI/ML application requests to perform data augmentation to enhance its training dataset.

* Step 1: AI application (oneM2M AE) sends request to the <dataAugmentation> resource
* Step 2: CSE stores received input to the <dataAugmentation> resource
* Step 3: Based on the given parameters, CSE tries to get the source image, apply the given data augmentation technique (e.g., resize), and generate target resources containing generated augment images

### -----------------------End of change 1-------------------------------------------