Service ID S00115

Location Italy, Remote

Data validation and processing

Provider service

Politecnico di Milano (POLIMI)

Link to content

https://www.agrifoodtef.eu/services/data-validation-and-processing

Type of Sector

Arable farming, Food processing, Greenhouse, Horticulture, Livestock farming, Tree Crops, Viticulture

Accepted type of products

Data, Design / Documentation

Type of service

Data analysis, Data augmentation

Description

Through this service, we offer customers the ability to validate and augment the data collected during the different testing phases. Examples of operations that this service can provide include quality checks (e.g., presence of all datastreams, presence of drop-outs, technical quality such as focus or exposure of camera images, agronomic significance of data); preparation for performance evaluation (e.g., identification of the images containing selected visual markers or specific plants, association between images from the system under test and the ground truth); data processing (e.g., transcoding) and augmentation (e.g., generation and incorporation of metadata). Data labelling (e.g., to enable training of AI models) is not included, as it is the object of a separate service (S00291).

How can the service help you

Generation of large amounts of data is a common byproduct of experimental activity on AI- or robotics-based systems. In order to become valuable, for instance, to be used to focus product development, raw data often requires validation (to verify their suitability for the task) and processing (to augment them in a way that prepares their optimal exploitation).

Such activities can require a very wide range of specialised expertise, comprising disparate items such as know-how about what can be expected by a given type of sensor in a specific experimental condition (e.g., to detect if the technical quality of sensor streams is subpar), agronomic know-how (e.g., to identify which of the plants appearing in a video stream correspond in their growth stage or disease status to the use case targeted by the customer), or the capability of writing custom transcoding software (e.g., to make the data compatible with a given software package).

This service provides the customer with access to an expert group of engineers and agronomists that possess, collectively, the whole range of capabilities to perform all of the above data operations and more. When needed, the service also provides the processing resources required for the execution of complex data processing tasks.

How the service will be delivered

This service description is intentionally generic. Every instance of this service is, in fact, customised to adapt it to the needs and requirements of the specific customer. The following is an example of a service instance.Example service: In the context of autonomously monitoring the hydric stress of grapevines, the customer has already collected hyperspectral and RGB images of Red Globe bunches at their premises. The customer asks for service S00115 (data validation and processing) in order to prepare the data for usage in AI training. Analysing this dataset, we notice that 10% of the provided frames are corrupted; therefore, we proceed to filter out the damaged frames so they do not negatively affect the inference model.

To enhance the quality of the hyperspectral data, we apply opportune noise reduction methods to the data. Additionally, let us assume the customer data were all collected in a single run at 9:00 AM. In this case, we would diversify the brightness of RGB images in post-processing to obtain a more heterogeneous and rich set that resembles different lighting conditions.

Service customisation

The duration of this service heavily depends on the type of processing that the customer requires, as well as on the amount of data to be processed. As a guideline, in typical cases the service may require 3-4 weeks.

The preparation phase of the service involves one or more interviews where the customer shares with AgrifoodTEF information about the data to be processed and the goals that they need to attain through such processing. Additional information (such as the technical specifications of the (sub)systems used to generate the data) may be requested to enable the activities of AgrifoodTEF; any confidential information will be shared under NDA.

The preliminary exchange of information is followed by the transmission to AgrifoodTEF of the data to be processed. Transmission can occur, for instance, by providing AgrifoodTEF with either a copy of the data or access to a repository where the data are stored. At the end of the service, the customer receives the processed data and a report detailing all issues detected in the original data (if any) and the impact that they had on the augmented data provided by the service.