



iImagine

Best practices and guideline for developers and providers of AI-based image analytics services

iImagine Deliverable 4.1

28/11/2022

Abstract

iImagine is a 36-month-long project to serve aquatic researchers with a portfolio of 'free at point of use' image datasets and high-performance image analysis tools empowered with Artificial Intelligence (AI). The iImagine services will build on an AI Platform that will be established by Q1 2023 and will allow transparent training, sharing, and serving of Machine learning (ML) and Deep Learning (DL) applications. This document provides good practice usage guides and pointers to documentation relating to the baseline technology of the iImagine AI platform, the DEEP-Hybrid-DataCloud services that have been developed and used in various EC initiatives in the past years. A more mature version of this document will be issued towards the end of the project.



**Funded by
the European Union**

iImagine receives funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101058625.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, which cannot be held responsible for them.

Document Description

D4.1 – Best practices and guideline for developers and providers of AI-based image analytics services

Document title			
Work Package number			
Due date	30/11/2022	Actual delivery date:	28/11/2022
Nature of document	[Report]	Version	1.0
Dissemination level	Public Confidential		
Lead Partner	CSIC		
Authors	Ignacio Heredia (CSIC), Valentin Kozlov (KIT)		
Reviewers	Dick Schaap (MARIS), Marco Rorro (EGI)		
Public link	10.5281/zenodo.7372358		
Project website	https://iImagine-ai.eu/		

Description

Abstract	
Keywords	Aquatic science, iMagine, AI platform, best practices

Revision History

Issue	Item	Comments	Author/Reviewer
V 0.1	Draft version	Table of Content	I.Heredia (CSIC), V.Kozlov (KIT)
V 0.2	Revised version	First full draft, revised internally by co-authors	I.Heredia (CSIC), V.Kozlov (KIT) / V.Tran (IISAS)
V 0.3	Revised version	Reviewed	D.Schaap (MARIS) M. Rorro (EGI)
V 0.4	Revised version	Updated after the review	V.Kozlov (KIT)
V 1.0	Submitted version		V.Kozlov (KIT)
V 1.1	Corrections		

Copyright and license info

This material by Parties of the iMagine Consortium is licensed under a [Creative Commons Attribution 4.0 International License](#).

Table of content

Contents

Introduction.....	5
Purpose of the document	5
Scope of the document.....	5
Structure of the document.....	5
The iMagine AI platform.....	6
How to access	6
Registration	7
User roles and workflows.....	7
The basic user	7
The intermediate user.....	8
The advanced user.....	8
Typical scenarios	8
Use an existing module	9
Train an AI module.....	9
Develop an AI module	10
Communication channels for users and providers	10
Glossary.....	11

Introduction

Purpose of the document

The iImagine image analysis services will build on the common 'iImagine AI Platform' that will be established by Q1 2023, and will allow transparent training, sharing, and serving of Machine learning (ML) and Deep Learning (DL) applications. The iImagine AI platform is built on the DEEP-Hybrid-DataCloud (hereinafter referred as DEEP) services that have been developed and been used in various EC initiatives in the past years. This document aims to accelerate the implementation and porting of the iImagine use cases to the future iImagine AI platform, by collecting all the documentations and good practice guides that are available about the DEEP system. The steps and practices that these documents introduce will only slightly change as they become part of the iImagine AI platform, therefore they can already serve as learning material for users and providers of the project.

The iImagine AI framework offers a portfolio of services for AI model development, training, and deployment, to be adopted by researchers in aquatic sciences. The collected practices and guidelines provide information on how to develop and operate AI-based image analytics services, which are essential to produce good outcomes in a productive manner.

Note: The DEEP services will provide the basis for the iImagine AI platform which will be central in the further development, testing, and deployment of the iImagine use cases. The planned iImagine AI platform is being configured and should be available in the 1st quarter of 2023; therefore, this guideline provides references to the existing DEEP platform.

Scope of the document

The document contains all the necessary information for iImagine users in order to start efficiently using the DEEP services. It overviews the services and workflows, providing pointers to the appropriate online documentation for further in-depth details. In this way, this document will be relevant all along the project, as the online documentation will be continuously updated to reflect the latest best practices, guidelines, and tutorials.

Structure of the document

The deliverable is organised as follows: the main components of the platform are listed with corresponding entry points and online documentation. Then targeted user roles and workflows are described referring to the platform components. This is followed by typical usage scenarios. In the end, the communication channels between users and the platform support are described.

The iImagine AI platform

The iImagine AI platform builds on the services provided by DEEP for artificial intelligence (AI), machine learning (ML), and deep learning (DL). Therefore, references in this guideline are pointing to available information and documentation about these services and workflows.

How to access

There are several services that are relevant for users:

- **Project's Homepage** - <https://deep-hybrid-datacloud.eu/>
A high level overview of the DEEP project.
- **Documentation** - <https://docs.deep-hybrid-datacloud.eu/>
This is the main source of knowledge on how to use the platform. Refer always to here in case of doubt. The documentation covers User documentation but also Technical notes about the platform itself.
- **YouTube channel** - https://www.youtube.com/playlist?list=PLJ9x9Zk1O-J_UZfNO2uWp2pFMmbwLvzXa
YouTube channel of the DEEP project with a number of video tutorials and examples.
- **Marketplace/Open Catalogue** - <https://marketplace.deep-hybrid-datacloud.eu/>
This is where users will typically search for AI modules developed by various communities, and find the relevant pointers to use them. If users develop new AI modules, they can share them on the Marketplace/Open Catalogue.
- **Dashboard** - <https://train.deep-hybrid-datacloud.eu/>
This is the place where users will go to request Cloud resources, including specific hardware (e.g. GPUs), to train an AI module. Access is restricted to authenticated users. The iImagine platform will provide unified access to the computing resources of the four cloud providers (CSIC, INCD, Walton, TUBITAK), members of the consortium.
- **DEEP IAM** - <https://iam.deep-hybrid-datacloud.eu/>
This is the authentication manager of the platform, where everyone should register to get access to the protected resources, like the Dashboard (see above).
- **DEEP-NextCloud** - <https://data-deep.a.incd.pt/>
This is the service that allows remotely storing the data and accessing them from inside your deployment. Access is restricted to authenticated users.
- **GitHub** - <https://github.com/deephdc>
This is where the code of all the AI modules and services behind the project can be stored.
- **DockerHub** - <https://hub.docker.com/u/deephdc/>
This is where the Docker images of the AI modules are stored.

- **Status of services** – <https://status.deep-hybrid-datacloud.eu/>
This is where you can check if a specific DEEP service might be down for some reason.

Note that for the aquatic science community, the iMagine platform will host dedicated instances of some of the services listed above, e.g. the Open Catalogue, Dashboard. If some of the URL addresses get modified during the iMagine project, users will be notified and any major lasting changes will be documented in the platform documentation. Please refer to the above links when those services are mentioned in the deliverable, unless stated otherwise.

Registration

Using advanced DEEP services, like training or storage, requires authentication. Please go to the DEEP-IAM web page and register for a DEEP-IAM account. Once your account is approved, you will be able to access authenticated services like the Dashboard.

User roles and workflows

The DEEP services are focused on three different types of users, depending on what they want to achieve. These roles are extensively described in the documentation¹, so only a summary overview is given in this guideline.

The basic user

This user wants to use AI modules that are already pre-trained and test them with their data, and therefore doesn't need to have any machine learning knowledge. For example, these users can take an already trained module for phytoplankton species classification that has been containerized, and use it to classify their own images of phytoplankton species.

DEEP can offer to this user:

- a catalogue full of ready-to-use AI modules to perform inference with your data
- an API to easily interact with the inference service
- solutions to run the inference in the Cloud or in your local resources
- the ability to develop complex topologies by composing different modules, thanks to the REST API available in every module

¹ <https://docs.deep-hybrid-datacloud.eu/en/latest/user/overview/user-roles.html>

The intermediate user

The intermediate users want to retrain an available module to perform the same task but fine tune it to their own data. They still do not need detailed knowledge on modelling of machine learning problems, but typically do need basic programming skills to prepare their own data into the appropriate format. Nevertheless, they can re-use the knowledge being captured in a trained network and adjust the network to their problem at hand by re-training the network on their own dataset. An example could be a user who takes the generic image classifier model and retrains it to perform phytoplankton classification.

DEEP can offer to this user:

- the ability to train an out-of-the-box AI module from the catalogue on user's specific dataset
- an API to easily interact with the model
- data storage resources to access your dataset (e.g. DEEP-Nextcloud)
- the ability to deploy the developed service on Cloud resources
- the ability to share the module with other users in the user's catalogue

The advanced user

The advanced users are the ones that develop their own ML/DL models and therefore need to be competent in the area of AI/ML/DL. This would be the case, for example, if apart from using the provided image classification module, users want to perform an object localization task, which is a fundamentally different problem. Therefore, they will design their own neural network architecture, potentially re-using parts of the code from other modules.

DEEP offers to these users:

- a ready-to-use environment with the main DL frameworks as a dockerized solution running on different types of hardware (CPUs, GPUs, etc),
- data storage resources to access the user's dataset (e.g. DEEP-Nextcloud, ...),
- the ability to deploy the developed AI module on Cloud resources,
- the ability to share the module with other users in the open catalogue and provide all relevant metadata information,
- the possibility to integrate your AI module with the API to enable easier user interaction or building further services on top of the AI module.

Typical scenarios

There are three typical scenarios that match with the three typical user roles. For each of the roles/scenarios, it is described how they are relevant for iImagine users.

Use an existing module

This scenario corresponds to the basic user who wants to use an existing AI module available in the Marketplace².

There are three possible workflows:

- to deploy and operate the available AI module on the user's own local resources (personal computer, cloud, kubernetes, ...) running the module as a Docker image;
- to use the AI module for inference via serverless services in the Cloud. This workflow is currently unavailable but will be made available during the course of the iImagine project on the iImagine cloud resources;
- to deploy the AI module on the cloud resources of DEEP via the Dashboard. However, this is not the intended use, since the cloud resources available through the Dashboard are primarily dedicated for training.

For the purposes of the iImagine project, the most relevant existing module might be the *Phytoplankton classification*³ module. Keep in mind that the DEEP platform role is to host the models, not to vouch for them meeting an arbitrary accuracy threshold. It is up to the user that shares the model with the community to provide details on the accuracy of the model.

Train an AI module

This scenario corresponds to an intermediate user who wants to retrain an existing module on their own dataset⁴.

There are two possible options:

- to retrain the module locally (i.e. on the user's own computer). As all the code is open-source the user can very well use their own resources for training,
- to train the module on the connected cloud resources via the Dashboard. This is what the DEEP platform was developed for. With this workflow, the user would
 - upload the dataset to DEEP-Nextcloud,
 - deploy the AI module the user wants to retrain via the Dashboard,
 - mount their Nextcloud folder inside the deployment,
 - retrain the module using the DEEPaaS API

² <https://docs.deep-hybrid-datacloud.eu/en/latest/user/index.html#use-a-model-basic-user>

³ <https://marketplace.deep-hybrid-datacloud.eu/modules/deep-oc-phytoplankton-classification-tf.html>

⁴ <https://docs.deep-hybrid-datacloud.eu/en/latest/user/index.html#train-a-model-intermediate-user>

A typical task that could be performed by an iImagine user would be to train the image classifier on a marine dataset, like what was done in the *Phytoplankton classification* module. Other image-related modules that could be useful to iImagine users are, for example, the *Image segmentation* module⁵ or the *Object detection* module⁶.

As a part of Best practices, it is planned to record and release the whole process describing how the *Phytoplankton classification* module was re-trained from the general *Image classification* module, as a paradigmatic example on how the platform could be used.

Develop an AI module

This scenario corresponds to advanced users that do not find an existing AI module according to their needs in the marketplace, and thus want to develop a new one from scratch⁷. This is probably the scenario which will fit the majority of the iImagine use cases. The workflow consists of using the DEEP Module Template to create a skeleton of the module code with the predefined structure, necessary packages, and other tricks that facilitate the development, training, and integration of the module in the marketplace. Then the user either adds already existing code to the skeleton or writes the code from scratch. Finally, the user creates a Pull Request on GitHub to integrate the new AI module in the Marketplace. Once the module is integrated in the Marketplace, refer to the *Train a module* section to see how to train it.

Communication channels for users and providers

For the purposes of the iImagine project, support should be asked in the mailing list of the user work package “AI technical integration and support”: iimagine-wp3@mailman.egi.eu. This will enable iImagine users, i.e. researchers in aquatic sciences, to profit from each other’s questions, which they may face themselves at some point. When questions might get too specific and are not covered by the general documentation of the platform, a dedicated meeting with the DEEP support team can be organised on-demand. The email list above can also be used for collecting the feedback and ideas about the best practices and guidelines. The DEEP services documentation is primarily hosted in GitHub⁸ and any suggestions on improvements can be submitted via GitHub issues or direct contributions via Pull Requests.

⁵ <https://marketplace.deep-hybrid-datacloud.eu/modules/deep-oc-semseg-vaihingen.html>

⁶ <https://marketplace.deep-hybrid-datacloud.eu/modules/deep-oc-obj-detect-pytorch.html>

⁷ <https://docs.deep-hybrid-datacloud.eu/en/latest/user/index.html#develop-a-model-advanced-user>

⁸ <https://github.com/deephdc/deep-docs>

To further foster communication, there planned dedicated webinars and workshops for users and providers. Once the iMagine platform is open for the broader community, a general support email list or a corresponding helpdesk will become available.

Glossary

Acronym	Description
AI	Artificial Intelligence
AI model	A complex algorithm that interprets data and makes decisions based on data. It has to be trained to obtain the most optimal set of model parameters, often called 'weights' in deep learning
API	Application Programming Interface
DEEP	DEEP-Hybrid-Datacloud platform
DEEP module	A software package, which includes an AI model, possible pre-processing, and DEEP API. A module is usually packed in a docker image. The REST API functionality allows to deploy a module as a service
DL	Deep Learning
IAM	Identity & Access Management
ML	Machine Learning
URL	Uniform Resource Locator