

The iMagine Al platform for image analysis – Opportunities for aquatic sciences

Gergely Sipos (EGI Foundation) On behalf of the iMagine consortium

EGI 2023 Conference 21/June/23, Poznan







Outline

- Project facts
- Technical architecture
- Lessons learned

Imaging services and applications

• Opportunity to work with iMagine



Marine environmental management and policy making



• 50% of the global oxygen production is produced by photosynthesis of marine algae. Oceans are home to the world's largest diversity of species and habitats. Annually 100 Mio tons of marine organisms are exploited as food source.

Relevant EU Directives and initiatives for aquatic domain, such as:

- Marine Strategy Framework Directive (MSFD)
- Water Framework Directive (WFD)
- European Green Deal
- Mission Starfish 2030 "Healthy oceans, seas, coastal and inland waters"
- United Nation's 2030 Agenda for Sustainable Development
- United Nation's Decade of Ocean • Science (2021-2030)



iMagine

Implementation requires knowledge

The implementation requires an increase of our overall knowledge, demanding more science and improved access to observation data and analytical processing.





Europe already has developed an impressive capability for aquatic environmental observation, data-handling and sharing, modelling and forecasting, second to none in the world. This builds upon national environmental observation and monitoring networks and programs, complemented with EU initiatives such as the Copernicus programme (CMEMS) and EMODnet, and European Research Infrastructures (RIs).

Project factsheet iMagine

OBJECTIVE:

To deploy, operate, validate, and promote a dedicated iMagine Al framework and platform, connected to EOSC and AI4EU, giving researchers in aquatic sciences open access to a diverse portfolio of AI based image analysis services and image repositories from multiple RIs, working on and of relevance to the overarching theme of 'Healthy oceans, seas, coastal and inland waters'.

- 36 months
- From Sept. 2022 until Aug. 2025
- €4.5 million EC funding
- 23 participants (19 beneficiaries + 4 affiliated partners)
- 18 service installations (Virtual Access)







4 national cloud compute centres and 5 AI/ML technology development institutes supports 14 research institutes for 12 Rls



EGI Conference 2023, P







Magine5 mature use cases

Aquatic Litter Drones: Aquatic Litter monitoring system using drones	DFKI (DE) + MARIS (NL) + OGS (IT)	Drone Survey - Level B
EcoTaxa pipeline: Taxonomic identification of zooplankton using Zooscan	Sorbonne Université (LOV+IMEV) (FR)	
Ecosystem monitoring at EMSO sites by video imagery	EMSO ERIC (IT) + UPC (ES) + Ifremer (FR) + MI (IE)	
Oil Spill Detection: Oil spill detection from satellite images	CMCC (IT) + OrbitalEOS (ES) + Uni. of Trento (IT)	
Taxonomic identification of phytoplankton using Flowcam images	VLIZ (BE)	60% oxygen on Earth!



Underwater Noise Identification: Underwater nois identification from acoustic recordings using spectrograms

Beach Monitoring: Posidonia oceanica berms and currents detection from beach monitoring system

Identification of diatoms in freshwater (unicellula microorganism)

se	VLIZ (BE)	Spectrogram
d rip- ms	SOCIB (ES)	<image/>
ar	UL-LIEC (FR) + CNRS-IRL2958 (FR) + SU-LOCEAN (FR)	





Operational iMagine platform with common AI development framework

Launch of 5 aquatic Al image analytics services, running operationally on the iMagine platform

3 Al-based imaging processing application pilots, 8 scientific image repositories

Objective 4. Capture and disseminate development and operational *Best Practices documentation*, best practices to imaging data and image analysis service providers interaction with EOSC and AI4EU platforms. + Training programme

Portfolio: operational services, image repositories, Best Practices, iMagine framework and platform









iMagine

Our technical architecture







- 5 production Al services
- 3 AI application prototypes
- ... for aquatic sciences

Kernel Tabs Settings Help			🚇 Alvaro Lopez Gai
t C 🖪 root@mesos-agent-gpu4: /×			
root8mesos-agent-gpu4:/srv# deep-start -do Last Modified Fri Apr 3 09:15:14 2020			Q. Search
a day ago NVIDIA-SMI 440.33.01 Driver Version: 440.33.01 CUDA Version: 10.2			
a day ago GPU Name Persistence-M Bus-Id Disp.A Volatile Uncorr. ECC	🏷 Model Trainable Inference	🏷 Model Trainable Inference	St Model Trainable Inference
a day ago Fan Temp Perf Pwr:Usage/Cap Memory-Usage GPU-Util Compute M.	DEEP OC Massive Online Data	DEEP OC Betinonathy Test	Train an image classifier
a day ago 0 Tesla V100-PCIE Off 00000000:00:06.0 Off 0	Streams	A Tensorflow model to classify	Train your own image classifier with
a day ago N/A 36C P0 24W / 250W 0MiB / 32510MiB 0% Default	Deep learning for proactive network	Retinopathy.	your custom dataset. It comes also
a day ago	monitoring and security protection.		pretrained on the 1K ImageNet class
seconds ago ++ Processes: GPU Memory			
a day ago GPU PID Type Process name Usage	St Model Trainable Inference	S Model Trainable Inference	S Model Trainable Inference
a day ago No running processes found	Conus species starsifier	Phytoplankton coocies	Sood species classifier
a day ago ++ [INFO] NVIDIA is present	Classify conus images among 70	classifier	Classify seeds images among 700K
a day ago [DEBUG] cpu: 'false', gpu: 'true', deepaas: 'true', jupyter: 'false', rclone: 'false', onedata:	true' species.	Classify phytoplankton images among	species.
a day ago [INFO] Checking ONECLIENT_ACCESS_TOKEN environment variable		60 classes.	
a day ago [INFO] Checking ONECLIENT_PROVIDER_HOST environment variable			
a day ago Connecting to provider 'cloud-90-147-75-163.cloud.ba.infn.it:443' using session ID: '33901943878	2632773'	Model Inference	S* Model Trainable Inference
a day ago Getting configuration Oneclient has been successfully mounted in '/mnt/onedata'.			
a day ago [ONEDATA] PID=1441	Speech keywords classifier	Body pose detection	Train an audio classifier
a day ago [DEPaas] deepaas-runopenwhisk-detectlisten-ip=0.0.0.0listen-port=31028	Irain a speech classifier to classify audio files between different keywords.	Detect body poses in images.	Irain your own audio classifier with your custom dataset. It comes also
a day ago ## ###			pretrained on the 527 AudioSet
a day ago ## ###### ##			classes.
a day ago .###### ###### ##############.			
#### ### ### // ### ## ## ## ## #### ##	* Development	ST Model Trainable	St Model Trainable Inference
Hybrid-DataCloud ##	DEED Development		
	DEEP Development	IF Benchmarks	Object Detection and





Generic, scalable platform for AI/ML applications

EGI federated cloud infrastructure OpenStack GPUs, CPUs, Storage in Spain, Portugal, Turkey, Ireland

1500 TB-months 132,000 GPU-hours 6,000,000 CPU-hours











iMagine Al Platform

🔿 iMagine				옵 Login - Re
Modules (marketplace)	Marketplace 注			Q Search modules
Other links 🖸	Train an image classifier	Object Detection and	Phytoplankton species	marine_species_seg
Identity and Access iMagine AI platform documentation Project page	Train your own image classifier with your custom dataset. It comes also pretrained on the 1K ImageNet classes.	Classification with Pytorch A trained Region Convolutional Neural Network (Faster RCNN) for object detection and classification.	classifier Classify phytoplankton images among 60 classes.	WIP Identification of marine species from EMSO Azores dee sea obervatory
	Trainable Inference Pre-trained	Trainable Inference Pre-trained	Trainable Inference Pre-trained	
	AI4OS Development Environment			
	This is a Docker image for developing new modules			
🔅 🚫 iMagine				
The iMagine platform dashboard is a service provided by CSIC, co-funded by iMagine				
Terms of use Privacy policy				





Common implementation approach



Continuous interaction with users, feedback & knowledge sharing

- Analysis of use cases using Persona-EPIC-User story approach
- Gap and Bottleneck analysis
- Initial Development roadmaps
- Initial collection of requirements for the platform

- Update of the roadmaps • Building up operation plans

Summing up lessons learnt for Best practices for

- image analysis in aquatic sciences
- Al service creation and operation

Development roadmap: <u>https://zenodo.org/record/7760413</u>





Lessons learnt... so far

- Labelling/annotating data: various tools are being assessed and used
- Building datasets for training and testing: the real world is noisy! \rightarrow Data quality check is of relevance and interest
- Working on metadata to improve FAIRness of data
- Selecting AI models for object detection, segmentation, further classification/taxonomy

 \rightarrow Monitoring of experimentation is helpful

- Engineering of end-to-end pipelines
- Regular knowledge exchange is important → Bi-weekly calls with all the use cases and AI platform experts
- Filling manpower gaps is hard...







.abelstud.i

Labelbox.com





The iMagine Open Call for Al-powered image analysis in aquatic sciences is NOW OPEN!

We offer:

- 10-month collaborative projects
- Support for
 - Al model training
 - Large-scale image analysis
 - Use of the iMagine AI platform to develop and train AI models
 - Access to cloud resources (GPUs, CPUs, storage) to store images and to scale up analysis workflows

train Al models ge) to store images Apply by 31 July





