DA4DTE: An Agentic System for Enhancing the Accessibility of Digital Twins of Earth

Myrto Tsokanaridou[1], Jakob Hackstein[2,3], Genc Hoxha[2,3], Sergios-Anestis Kefalidis[1], Konstantinos Plas[1,6], Begum Demir[2,3], ***Manolis Koubarakis[1,6]****, Marco Corsi[4], Cristian Leoni[4], Giorgio Pasquali[4], Simone Tilia[4], Chiara Pratola[4], Nicolas Longepe[5]*

[1]Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece
[2]Technische Universitat Berlin, Germany
[3]BIFOLD - Berlin Institute for the Foundations of Learning and Data, Germany
[4]e-GEOS S.p.A., Italy
[5]Φ-lab, ESA ESRIN, Frascati, Italy
[6]Archimedes, Athena Research Center, Greece

Why a Digital Assistant for Digital Twins of Earth?

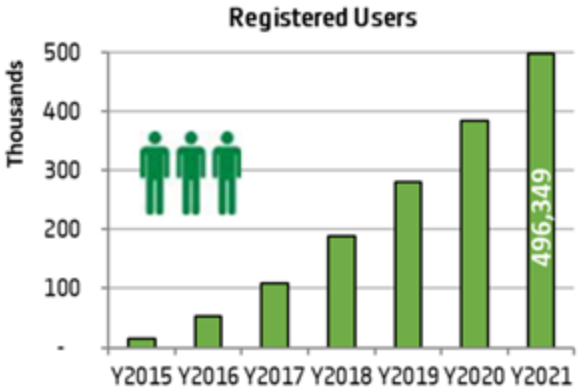
Advances in satellite technology have led to regular, frequent, and high-resolution monitoring of Earth at global scale, providing an unprecedented amount of Earth observation data.

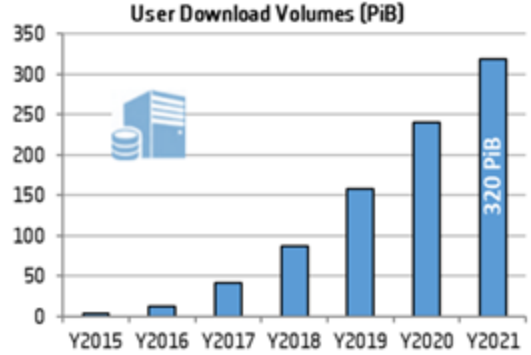
Through the Copernicus programme, Sentinel satellites reach the scale of more than 10 TB data per day.

Availability of a huge amount of satellite data has led to:

1. **Increased demand for EO data** from "laymen" such as journalists, business managers, university researchers, students and those that are simply curious.

1. The implementation of **increasingly complex catalogues** in order to manage an ever-expanding amount of data in an even shorter time.





Charts taken from Copernicus Sentinel Data Access Annual Report 2021

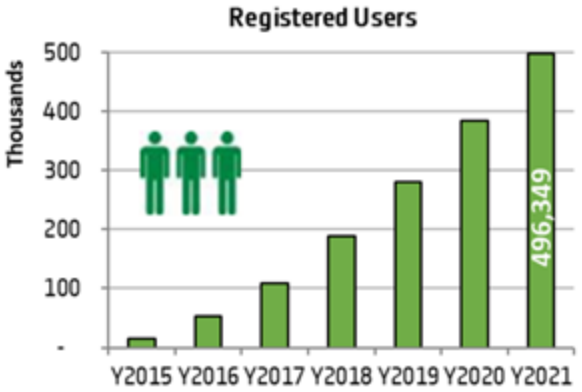
Why a Digital Assistant for Digital Twins of Earth?

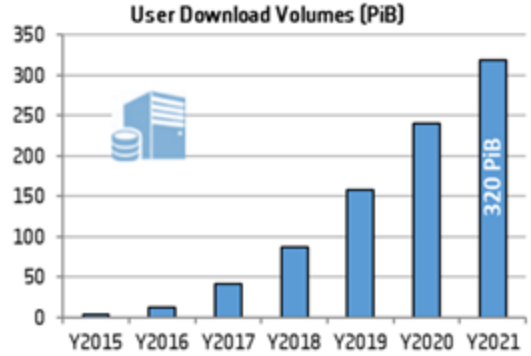
We need a platform that simplifies search and retrieval from these massive archives and that also allows non-expert users to fully benefit from EO data.

The main objective of the project was to develop a **first generation, AI-powered "Digital Assistant" interface to EO data** by harnessing advances in and convergence of diverse digital technologies such as for example **NLP, CV, KG (knowledge graphs) and content-based retrieval**.

We introduce the ***Digital Assistant for Digital Twins of Earth*** (DA4DTE), an AI-powered multiagent system designed to facilitate seamless interaction with EO datasets:

- **Simplifies the use of EO catalogues**, allowing queries in natural language in order to easily retrieve data
- **Facilitates the use of additional value** that EO data sources could bring, simply by asking the platform to extract information from data





Charts taken from Copernicus Sentinel Data Access Annual Report 2021

Consortium presentation

INDUSTRIAL PARTNER

e-geos
AN ASI / TELESPAZIO COMPANY

- **Industrial competences for the integration, deployment and demonstration** of the precursor **Digital Assistant**.

SCIENTIFIC PARTNERS

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Technische Universität Berlin

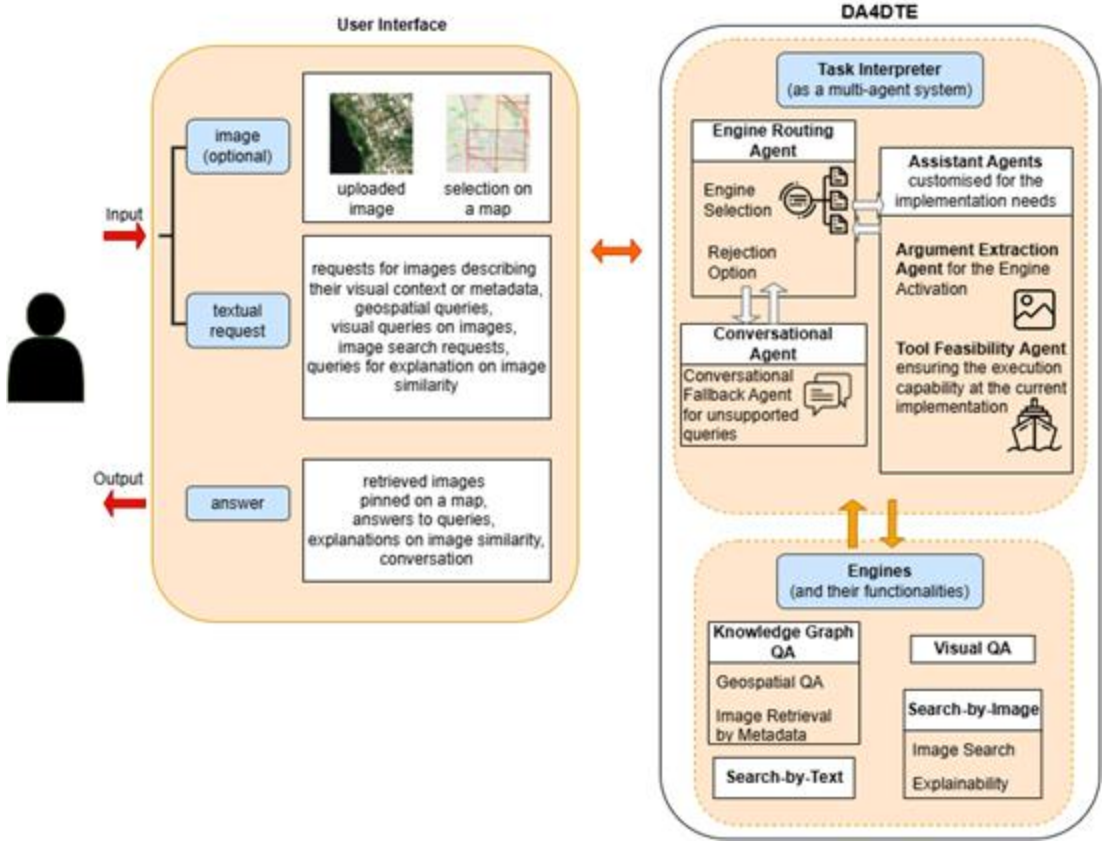
- **Scientific competences on the development of "Content-Based" Engines** exploiting NLP, CV, KG, CBIR and other techniques to develop the Search Engines.
- **Scientific network and credibility to disseminate the results of the work** in the broader community of NLP, CV, KG, CBIR scientists and practitioners, inside and outside the Earth Observation domain.

Architecture

Agentic Orchestration

The ***Digital Assistant*** consists of:

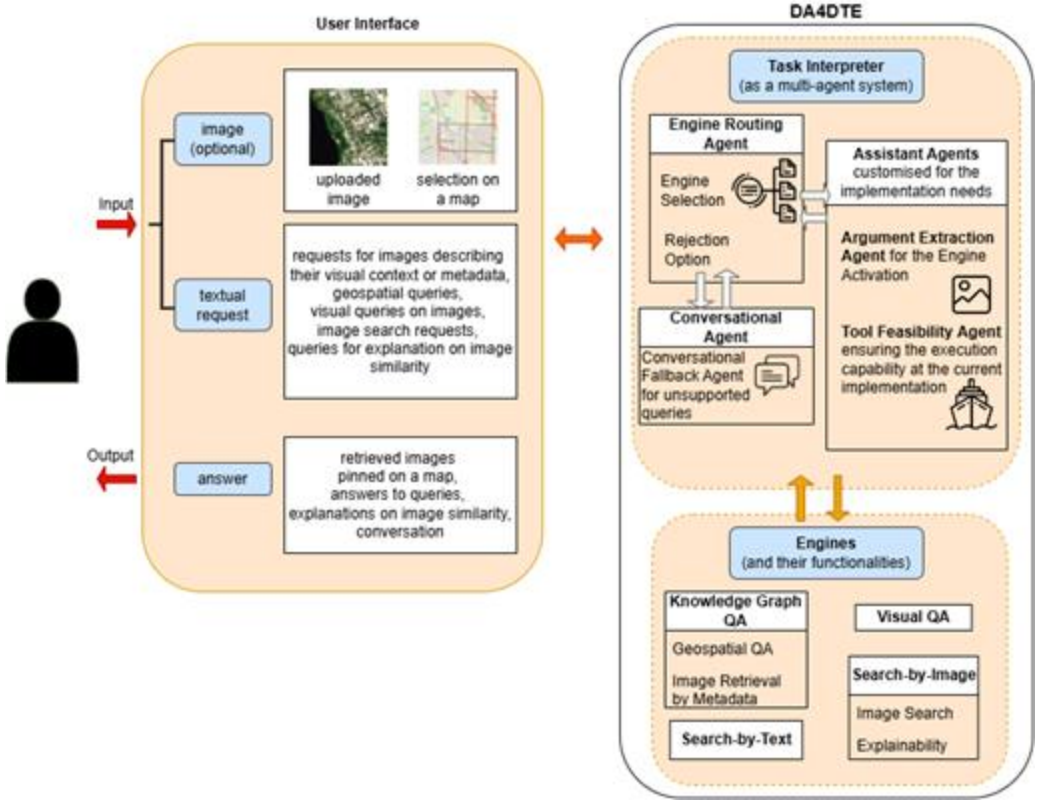
- ❖ **Front-end interfaces:** Web UI, API
- ❖ **Agentic Core:** Responsible for task orchestration and conversing with the user.
- ❖ **Content extraction engines:** Specialized tools for retrieving and analyzing images.
- ❖ **Auxiliary modules:** Knowledge sources and the necessary infrastructure for deploying and running the system.



Agentic Orchestration

The ***Agentic Core*** consists of Core agents and Assistant agents:

- ❖ **Engine Routing Agent** (Core): Zero-shot prompted LLM. Selects the engine to activate. Rejects requests that fall outside the scope of the available engines.
- ❖ **Conversational Agent** (Core): Designed to handle general, ambiguous, or out-of-domain queries. Prompted not to respond to irrelevant requests.
- ❖ **Argument extraction Agent** (Assistant): Extract key parameters required by specific tools
- ❖ **Tool Feasibility Agent** (Assistant): Responsible for validating whether a requested operation is feasible under current system capabilities

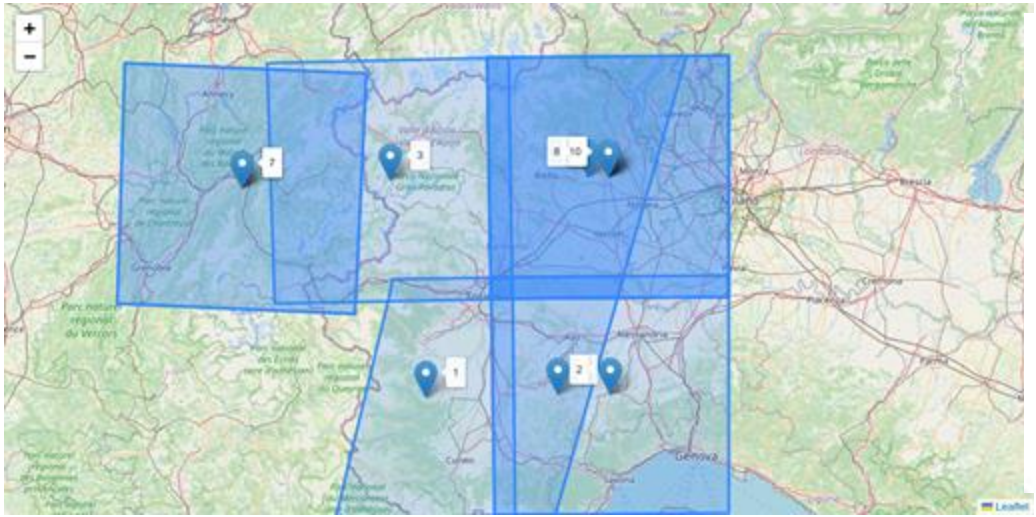


Content-based Extractor Engines

Knowledge Graph Question Answering Engine (TerraQ)

The ***Digital Assistant*** must be able to address requests for images that satisfy **spatial**, **temporal** and **metadata-related** constraints:

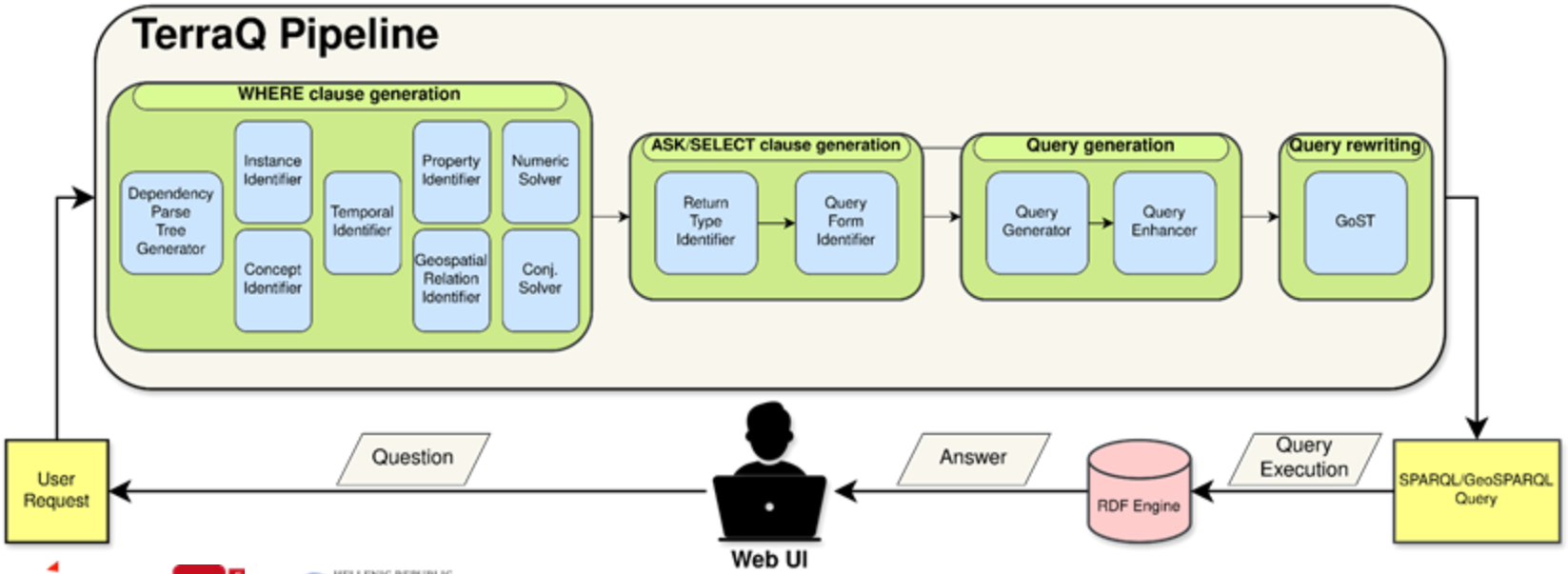
*Find 10 images of **Piedmont** with **cloud coverage under 20%** and **more than 50% vegetation**, taken in **August 2022**.*



Knowledge Graph Question Answering Engine

TerraQ is a spatiotemporal question-answering engine for satellite image archives that targets a ***purpose-built knowledge graph*** which contains image metadata and geospatial information for administrative divisions and natural features.

- TerraQ works by **translating** natural language requests into **GeoSPARQL** queries (Text-to-GeoSPARQL).
- Implemented as a **component pipeline** that combines classical NLP techniques (**fast**) with LLMs (**accurate**).



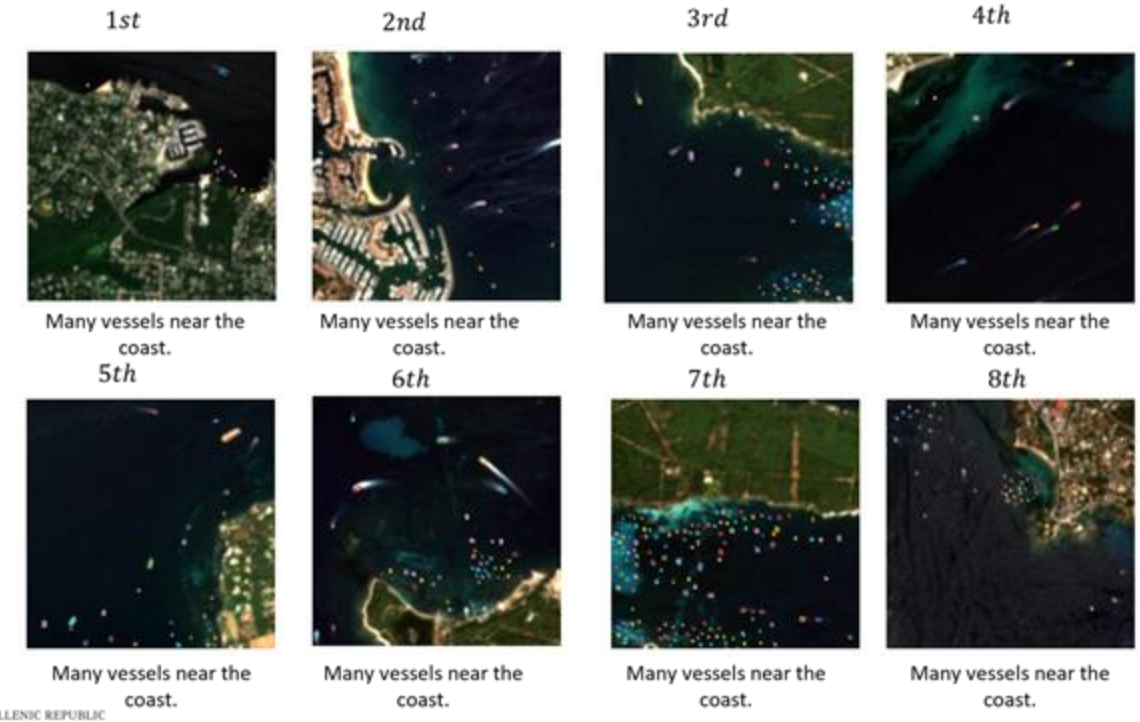
Sergios-Anestis Kefalidis, Konstantinos Plas, Manolis Koubarakis: *TerraQ: Spatiotemporal Question-Answering on Satellite Image Archives*. IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2025).

Search-by-Text Engine

The ***Digital Assistant*** must be able to retrieve images based on descriptions that cannot be answered by metadata :

Query Text

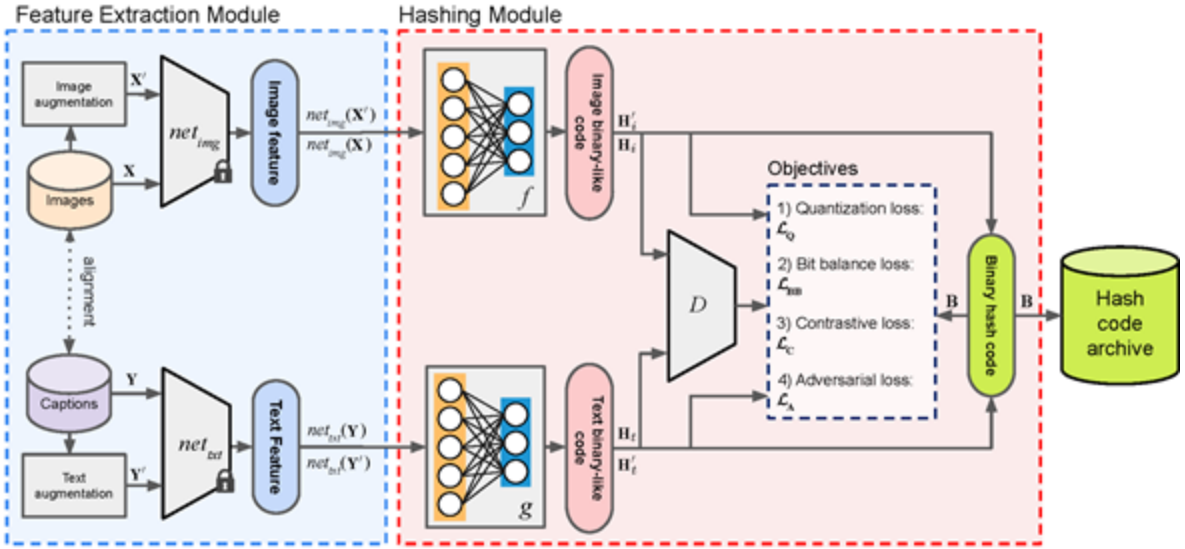
Many vessels near the coast.



Search-by-Text Engine

We employ the ***DUCH*** *(deep unsupervised cross-modal contrastive hashing)* method, which provides cross-modal retrieval between remote sensing images and text sentences:

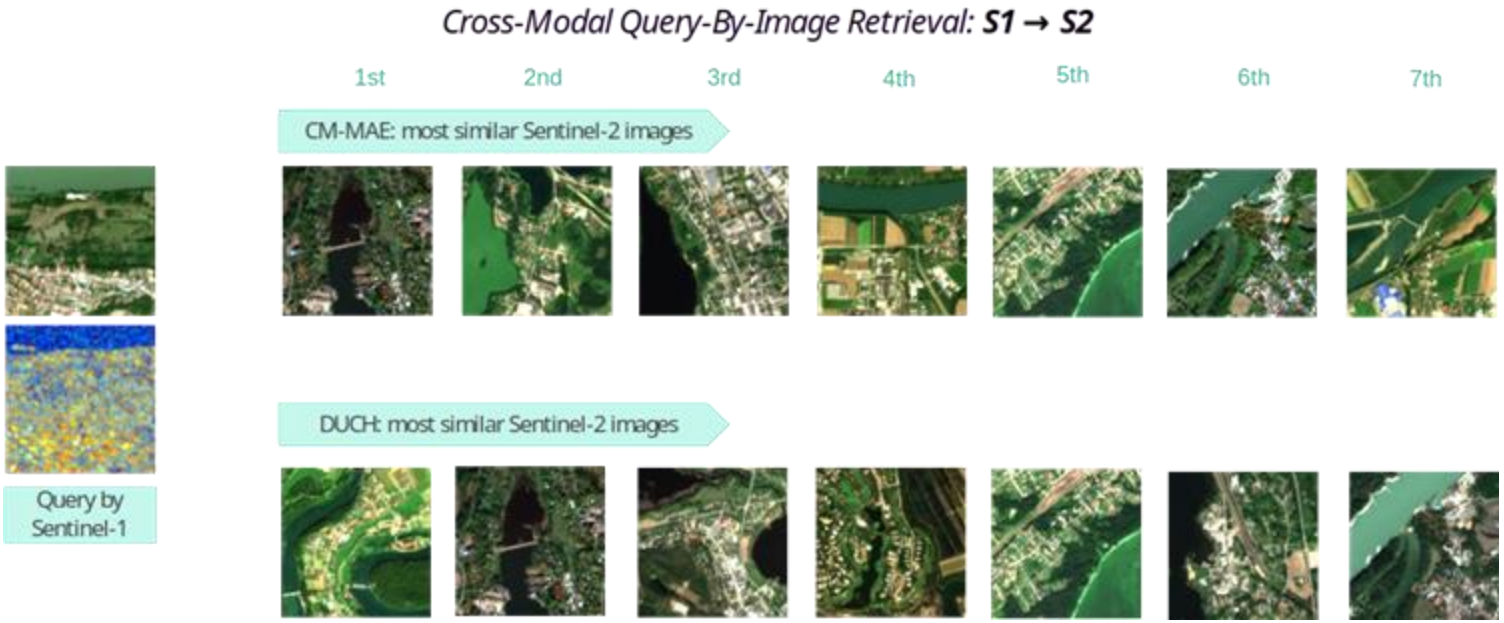
I. The feature extraction module produces deep semantic representations for image and text modalities
II. The hashing module that generates binary representations from the input extracted deep features
III. The hamming distance between images in the archive and the generated binary representation is used to evaluate similarity



G. Mikriukov, M. Ravanbakhsh, B. Demir, "*Unsupervised Contrastive Hashing for Cross-Modal Retrieval in Remote Sensing*," IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP).

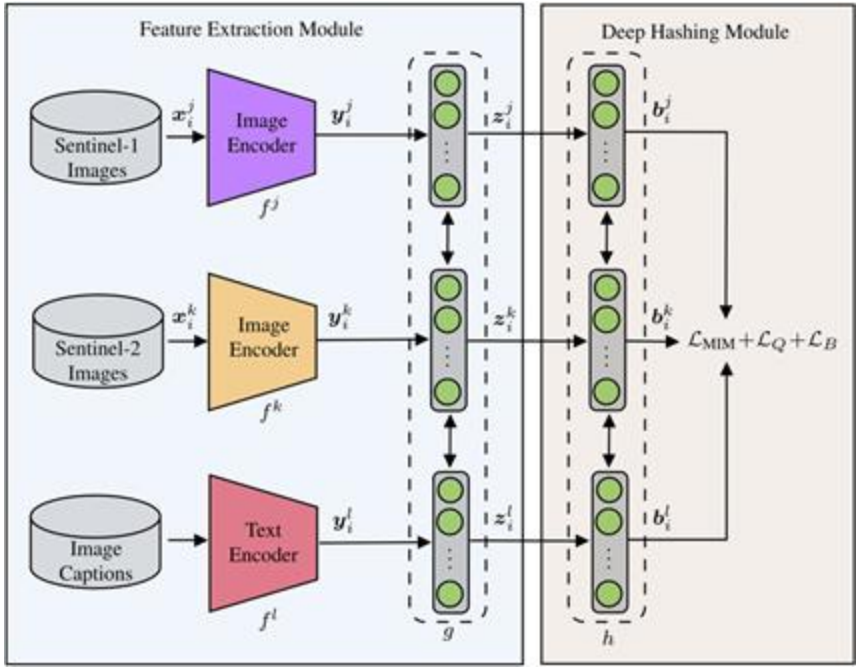
Search-by-Image Engine

The ***Digital Assistant*** must be able to retrieve images similar to a selected image (uni-modal and cross-modal):



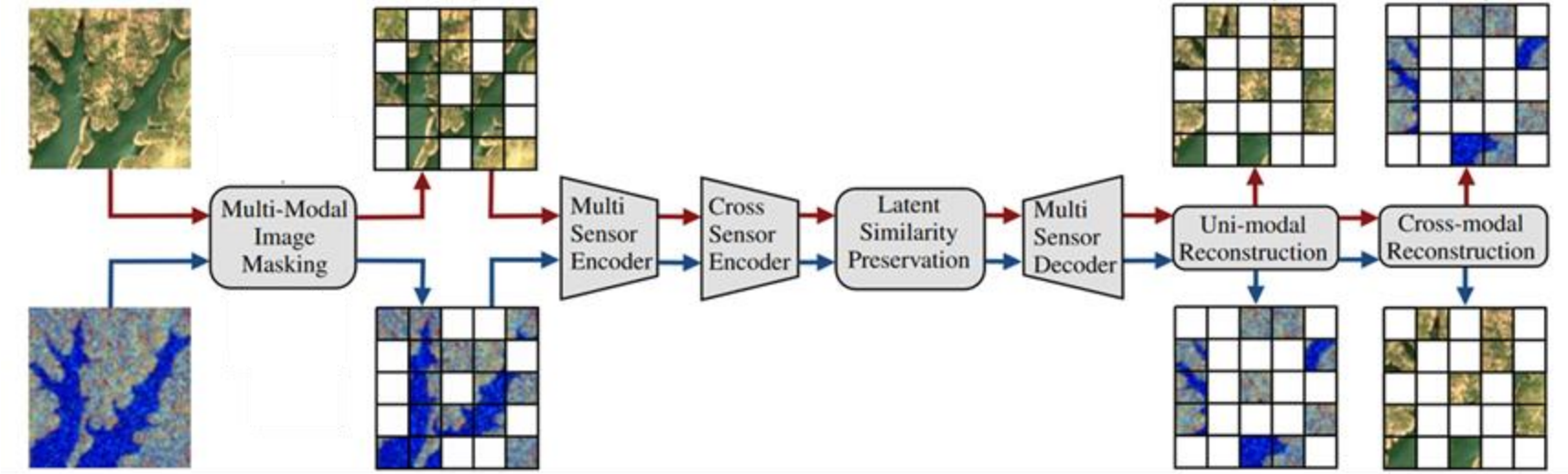
Search-by-Image Engine

We adapted the deep unsupervised cross-modal contrastive hashing (***DUCH***) method.



Search-by-Image Engine

We also developed the cross modal masked autoencoder (***CMMAE***) method.



Jakob Hackstein, Gencer Sumbul, Kai Norman Clasen, Begüm Demir: *Exploring Masked Autoencoders for Sensor-Agnostic Image Retrieval in Remote Sensing*. IEEE Trans. Geosci. Remote. Sens. 63: 1-14 (2025)

Visual Question Answering Engine

The ***Digital Assistant*** must be able to analyze images and answer questions about their contents.

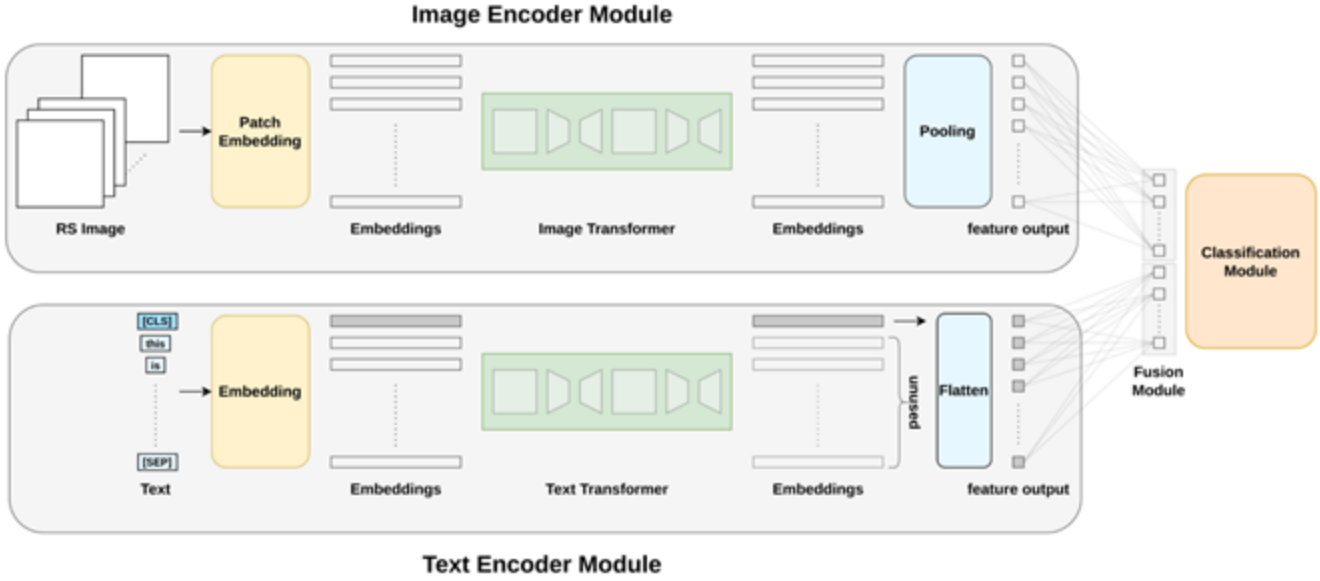
Question: Is there an agricultural area in this scene?
Answer: No

Question: Is this a rural area?
Answer: Yes

Visual Question Answering Engine

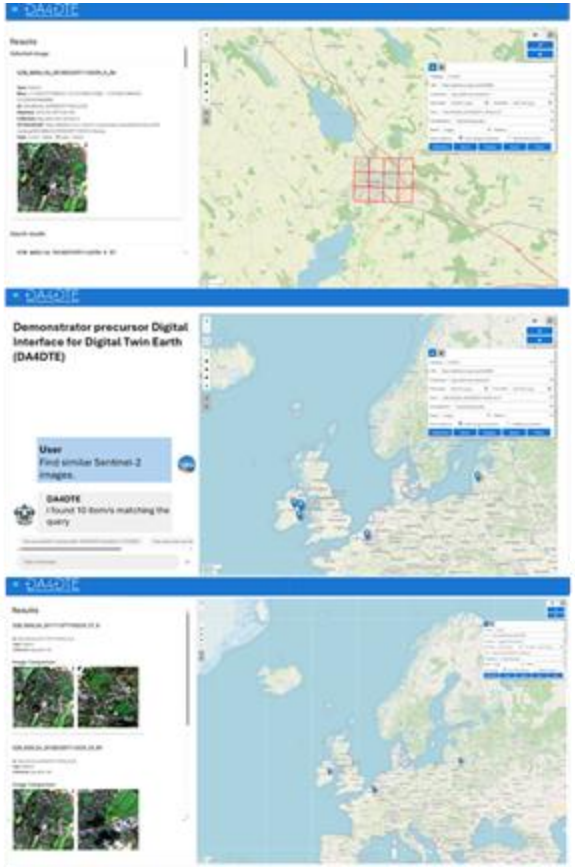
The ***LiT-4-RSVQA*** architecture provides **accurate** VQA results while significantly **reducing the computational requirements**:

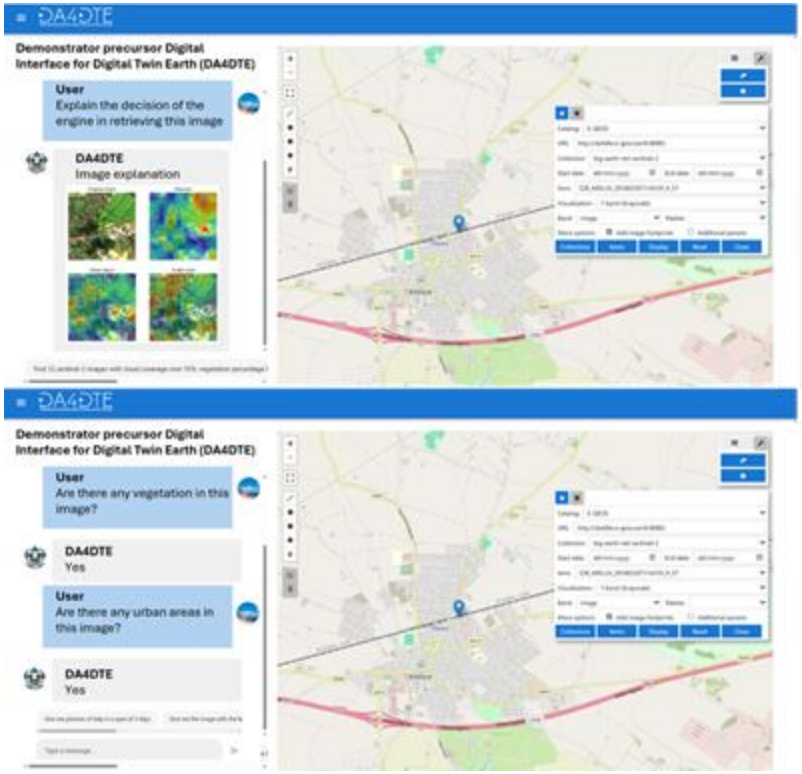
I. We use transformers to get representations (i.e., embeddings) for the question and image.
II. We project the embeddings in a common space.
III. We use the fused embeddings as input in a neural classification module.



Leonard W. Hackel, Kai Norman Clasen, Mahdyar Ravanbakhsh, Begüm Demir: *LIT-4-RSVQA: Lightweight Transformer-Based Visual Question Answering in Remote Sensing*. IGARSS 2023.

User Interface


Demonstrator precursor Digital Interface for Digital Twin Earth (DA4DTE)
User
Find similar Sentinel-2 images.
DA4DTE
I found 10 items matching the query


Demonstrator precursor Digital Interface for Digital Twin Earth (DA4DTE)
User
Explain the decision of the engine in retrieving this image
DA4DTE
Image explanation
Demonstrator precursor Digital Interface for Digital Twin Earth (DA4DTE)
User
Are there any vegetation in this image?
DA4DTE
Yes
User
Are there any urban areas in this image?
DA4DTE
Yes

Thank you for your attention!