



INTER IIT
TECH MEET 14.0

MID PREP



GeoNLI: Natural Language
Interpretation of Satellite Imagery

Introduction

We are in an era of unprecedented Earth observation. Every day, a large fleet of satellites captures petabytes of high-resolution imagery, documenting our planet's changes in remarkable detail. This data is critical for solving some of humanity's most pressing challenges. However, extracting meaningful insights requires specialized domain expertise, access to costly software, and often, significant programming skills. This high barrier to entry creates a critical bottleneck, leaving decision-makers, planners, and scientists dependent on a small pool of GIS experts. The result is a significant delay between data collection and actionable intelligence. This raises the question: can Artificial Intelligence simplify access to actionable intelligence through a natural language interface?

About Us

The Space Applications Centre (SAC) is a major research and development centre of the Indian Space Research Organisation (ISRO), situated at Ahmedabad. Its primary function is the design and development of space-borne and air-borne instruments/payloads for ISRO missions, alongside the development and operationalization of space technology applications for societal benefit. These developments and applications encompass critical domains such as communication, broadcasting, navigation, disaster monitoring, meteorology, oceanography, environmental monitoring, and natural resource surveying using microwave remote sensing, optical remote sensing, communication and navigation payloads. Furthermore, the Centre plays a pivotal role in India's scientific and planetary missions, including Chandrayaan-1, Chandrayaan-2, Chandrayaan-3 Landing mission, Aditya-L1, Mars Orbiter Mission and upcoming missions like Chandrayaan-4, Venus Orbiter Mission.

SAC's technical expertise also includes developing optical and microwave signal and image processing software and Geographic Information System (GIS) applications for ISRO's Earth Observation (EO) programme. These innovations support diverse fields, including Geosciences, Agriculture, Climate Change, Oceanography, and Atmospheric Science. To achieve its objectives, SAC actively collaborates with industry, academia, and national and international institutions.

Problem Statement Description

“Is the picture really worth a thousand words?”

Your challenge is to design and build a functional prototype that empowers non-expert users to interpret and analyze satellite imagery using the most intuitive interface of all: natural language.

The platform will accept a processed satellite image and be able to do the following key tasks:

- **Captioning:** A description of the image that clearly summarizes all the key information in the satellite image accurately.
- **Grounding:** Localize objects within the satellite image based on the natural language query. The outcome will be oriented bounding boxes overlaid on image.
- **Visual Question Answering:** Question answers that address the geometric and semantic attributes of the objects and features in the given image.

This problem statement seeks innovation over and above the existing methods to be integrated into the solution to improve:

- **Details:** Describe and measure the local and global attributes of small and large objects in the satellite image accurately.
- **Efficiency:** Analyse images up to 2k x 2k size quickly and efficiently.
- **Scale resilience:** Be able to handle sampling from 0.5m/pixel to 10m/pixel imagery efficiently.
- **Multimodality (Bonus!):** These add to innovation points if demonstrated. Can your solution work with SAR or Infrared Imagery? What about false color composites for optical images?

Methodology

1. Primary Solution format:

- a. A hosted website on the internet to be shared with anonymous URL, that has an interface to upload satellite images (L1/L2 processed, converted to 0-255 .png/.jpg format) and execute language based queries.
- b. Evaluation mode: Added support for query in evaluation mode (details will be shared in the company session soon)

2. The teams can utilize existing LLMs (Large-Language Models) and Vision-Language Models (VLMs) with proper credits to the respective authors. The additional improvements in accuracy using post-processing or fine-tuning (classical/deep learning) will likely improve the team-scores on innovation as well as evaluated metrics.

Evaluation

- 1. Core Idea (15%):** The solution idea at the back-end model execution (**5%**) and most importantly the innovation factors. This means not directly offering the existing solutions from literature but improving upon them to address accuracy, efficiency and resilience to adverse inputs. (**10%**).
- 2. Solution availability with complete codebase (15%):** The availability of the solution as a web-based platform with good design and user experience. Ability to work with larger images and intuitiveness of the interface. No glitches! This will be manually tested by judges by directly interacting with the platform interface. Code for front-end and back-end models should be submitted in the solution apart from the hosting of the website.
- 3. Objective Evaluation (50%):** A dataset consisting of the following category of images from will be used for evaluation through the hosted websites from the respective teams during the final evaluation:
 - a. Expert-captioned dataset including from ISRO's optical data sensors with natural color composite (R,G,B) bands. (Criteria: BLEU-Score with respect to expert annotations)
 - b. Satellite images with query text and a set of oriented bounding boxes as answers for that query. (Criteria: Intersection-over-Union @ 0.7)
 - c. VQA subset: A set of images with 3 challenge questions each (1 binary with yes/no answer, 1 numeric answer (always float) and 1 object attribute related (string)). (Criteria: as discussed in [1])
- 4. Documentation (10%):** The final report must be brief but to-the point description. We seek a 8 page (max) double column conference paper format report covering all relevant sections.

5. Presentation (10%): Impress with a neat summary of your work during presentation, with emphasis on innovation and key contributions.

Teams are encouraged to build the solutions with open datasets and test thoroughly for various types of input conditions. To score higher marks, teams will need to emphasize innovation beyond integrating the existing baselines.

Constraints and Rules:

1. The final solutions should not use direct commercial APIs/access or readymade solutions to achieve the objective.
2. All solutions must be anonymous in all aspects. The websites must use team numbers and there should not be design/text hints pointing to the name of the institute from which the team belongs.

Resources

- A thorough understanding of the problem, datasets and expected categories, evaluation criterion and the baseline solutions is available for reference at: <https://vrsbench.github.io/>.
- Evaluation mode details will be discussed and provided during the company session.