Code-based AI assistant for EO tasks

Post-doctoral position

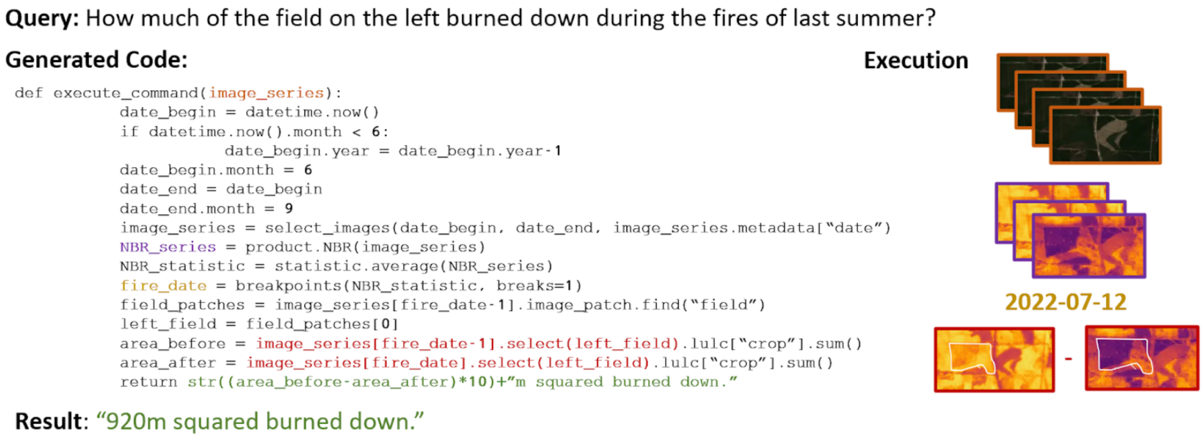
General information

- Keywords: Computer vision, Natural Language Processing, Deep Learning, Remote sensing, Multi-modality
- Duration of the post-doc: 12 months
- Institute: Université de Paris, Laboratoire d'Informatique Paris Descartes (LIPADE), team Systèmes Intelligents de Perception
- Location: 45 rue des Saints-Pères, 75006 Paris
- Supervisor: Sylvain Lobry
- Application: please send an email to sylvain.lobry "at" u-paris "dot" fr with the subject "[Postdoc IC-EO] FirstName LastName" containing:
  - updated CV;
  - cover letter;
  - contact information of a teacher/supervisor willing to write a recommendation letter.
- The position is open until filled, and no longer that the end of 2024.

Introduction

In recent years, remote sensing images have become more available than ever thanks to important efforts coming from the public and private sectors. For instance, the European Union's Copernicus program provides free access to Synthetic Aperture Radar (SAR) and multi-spectral data. In addition to governmental initiatives, companies (e.g. Planet Labs) also provide very-high resolution images on a global scale on a daily basis. Remote sensing images contain information which is already used, among others, to track climate change, improve security and to understand and manage the environment. Exploiting the different levels of information provided by the wide range of remote sensing modalities is an active field of research. Multi-modality is used in many remote sensing applications [1]. However, the interpretation of remote sensing data is generally performed by experts and often involves manual processing. With the increasing amount of data, the manual interpretation becomes a limiting factor impacting the delay at which information is extracted, but also the domains in which such data can be used. For specific applications, the remote sensing community has been developing ad hoc automatic methods. As such, these works can only address either general applications (e.g. pollution monitoring) or ones with direct financial interest. We argue that the information contained in remote sensing images can be of interest to a much larger public: journalists could retrieve such data to understand, follow and report on wars and the effects of climate change or local governments could use this data in their decision process and studies. While the data is present, the general audience do not always have either the technical knowledge to extract the information of interest or the capacity to fund research to do so. Enabling information extraction from remote sensing data through a non-technical and common interface would be a way to allow the general audience to directly benefit from this data.

The objective of the IC-EO project is to propose a conversational assistant that can interact with EO data, such as images. In this project, we will take advantage of the fact that EO tools can be interfaced via code. This will allow us to propose a model that can integrate current state of the art research in an extensible and straightforward manner. The objective of the IC-EO assistant is to be able to answer such questions as: "How much of the field on the left of this area burned down during the fires last summer?", and "How many buildings are less than 5 minutes walk from the park?", without restriction to specific formats. Uniquely, the answer to such questions will be code-based and hence can be reviewed and made comprehensible, as opposed to the black-box nature of current approaches. The model can be decomposed in two parts: first, we propose an application programming interface (API) that encompasses elementary tasks, such as classification, object detection, visual grounding and semantic segmentation. Second, we will train a large-language model to use this API, given as a context, to predict a code-based solution to the user's query. A sample of the execution of the assistant is shown in Figure 1.



Query: How much of the field on the left burned down during the fires of last summer?

Generated Code:

```
def execute_command(image_series):
            date_begin = datetime.now()
            if datetime.now().month < 6:
                        date_begin.year = date_begin.year-1
            date_begin.month = 6
            date_end = date_begin
            date_end.month = 9
            image_series = select_images(date_begin, date_end, image_series.metadata["date")
            NBR_series = product.NBR(image_series)
            NBR_statistic = statistic.average(NBR_series)
            fire_date = breakpoints(NBR_statistic, breaks=1)
            field_patches = image_series[fire_date-1].image_patch.find("field")
            left_field = field_patches[0]
            area_before = image_series[fire_date-1].select(left_field).lulc["crop"].sum()
            area_after = image_series[fire_date].select(left_field).lulc["crop"].sum()
            return str((area_before-area_after)*10)+"m squared burned down."
```

Result: "920m squared burned down."

Figure 1: Sample of the execution of the assistant for a query *How much of the field on the left burned down during the fires of last summer?*

IC-EO is an ESA (European Space Agency) project led by askEarth and Université Paris Cité. For a maximal impact, our dissemination strategy for this project includes published research, open sourcing developed models, and a demonstrator accessible online

Background

Interactions between textual features and images is a rising topic in the machine learning and computer vision communities. In particular, these interactions are essential components of tasks such as image captioning (IC) [2], image querying (IQ) [3] or Visual Question Answering (VQA) [4]. These tasks are particularly relevant when used with remote sensing data. Indeed, image querying has been a task of interest in the remote sensing community since the creation of massive remote sensing images databases as a way to explore them through natural language [5]. On the other hand, VQA has only been recently proposed in the remote sensing community [6] and has been identified as one of 6 potential game-changers in the field of Artificial Intelligence for Earth Science [7]. It aims at answering in English to a question (in English as well) about a remote sensing image. Since the introduction of this task in the remote sensing community in 2019, numerous research works have explored the construction of large databases for the training of supervised models [6, 8, 9] as well as the models themselves [6, 10, 11, 12]. However, these solutions often act as black-boxes, which prevent from reviewing the predicted answers. This can be particularly problematic when the extracted information needs to be used for making decisions. In this project, we propose to develop a methodology in the line of ViperGPT [13], which decomposes the answer retrieval process in small tasks that can be easily reviewed.

Objectives

The selected candidate will first review the state of the art for elementary tasks (e.g. classification, object detection, visual grounding, semantic segmentation) in order to create the API. In this part, the objective is to re-use existing open models. In addition, based on the candidate's interests, it is expected to work on a specific task to propose new methodologies. The candidate will then work in collaboration with askEarth to develop the large language model (LLM) fine-tunning for our specific assistant. In addition, we aim to study means for evaluating (quantitatively and qualitatively) the final model. This project is done in collaboration with askEarth. As such, the candidate will participate to a monthly meeting to share their progresses and propose feedback on the developments made by askEarth.

Desired background for the candidate

We are looking for a recently graduated PhD. The ideal candidate would have a theoretical background in computer vision and be proficient in programming with Python. Knowledge in geographic information sciences and natural language processing is a plus.

Bibliography

[1] Mauro Dalla Mura et al. "Challenges and opportunities of multimodality and data fusion in remote sensing". In: *Proceedings of the IEEE* 103.9 (2015), pp. 1585–1601.

[2] Quanzeng You et al. "Image captioning with semantic attention". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4651–4659.

[3] Huafeng Wang et al. "Deep learning for image retrieval: What works and what doesn't". In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE. 2015, pp. 1576–1583.

[4] Stanislaw Antol et al. "Vqa: Visual question answering". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2425–2433.

[5] Klaus Seidel et al. "Query by image content from remote sensing archives". In: *IGARSS'98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings.(Cat. No. 98CH36174)*. Vol. 1. IEEE. 1998, pp. 393–396.

[6] Sylvain Lobry et al. "RSVQA: Visual question answering for remote sensing data". In: *IEEE Transactions on Geoscience and Remote Sensing* 58.12 (2020), pp. 8555–8566.

[7] Devis Tuia et al. "Toward a Collective Agenda on AI for Earth Science Data Analysis". In: *IEEE Geoscience and Remote Sensing Magazine* 9.2 (2021), pp. 88–104.

[8] Sylvain Lobry, Begüm Demir, and Devis Tuia. "RSVQA Meets Bigearthnet: A New, Large-Scale, Visual Question Answering Dataset for Remote Sensing". In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE. 2021, pp. 1218–1221.

[9] Maryam Rahnemoonfar et al. "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding". In: *IEEE Access* 9 (2021), pp. 89644–89654.

[10] Christel Chappuis et al. "How to find a good image-text embedding for remote sensing visual question answering?" In: *MACLEAN Workshop at ECML*. 2021.

[11] Xiangtao Zheng et al. "Mutual Attention Inception Network for Remote Sensing Visual Question Answering". In: *IEEE Transactions on Geoscience and Remote Sensing* (2021).

[12] Sylvain Lobry et al. "Better Generic Objects Counting When Asking Questions to Images: A Multitask Approach for Remote Sensing Visual Question Answering". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2020, pp. 1021–1027.

[13] Dıdac Surıs, Sachit Menon, and Carl Vondrick. "Vipergpt: Visual inference via python execution for reasoning". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 11888–11898.