Visual Question-Answering for Thematic Maps

Eftychia Koukouraki
Institute for Geoinformatics, University of Münster, Germany

Date of submission: 10.11.2024

Required funding: 1.0 full-time equivalent for 6 months

Abstract

Visual question-answering (QA) helps users interpret complex visual information, making it easier and faster to gain insights also from maps and geospatial data in a variety of contexts. This project aims to create an open dataset tailored for thematic map-based QA systems, accompanied by a baseline model to demonstrate its usage. By compiling map images annotated with question-answer pairs, the dataset will enable Artificial Intelligence (AI) models to extract and interpret geographic and information from maps. The deliverables will include a curated dataset, a baseline model, documentation, and an evaluation report, all of which will be released under a permissive license to support further research on the topic.

I. *Introduction*

In the domain of Earth System Sciences, geographical maps are often used to visualize data, communicate research results and describe phenomena. Reading maps requires sophisticated reasoning, as it involves combining multiple visual variables in order to extract spatial information and draw insights. The results of our recent research show that content is very important to map readers when interacting with maps [1]. Question-answering (QA) serves as a method to extract and evaluate map content [5].

Recent advances in computer vision and large language models have extended what we have been able to achieve in terms of assisting human reasoning with technological tools. The computer vision task of teaching machines to comprehend the content of a picture and to answer questions about it in natural language is called visual question-answering (VQA). VQA applications are generally important for visually impaired people as they can significantly assist in the interpretation and description of visual content [2]. Moreover VQA systems can help overcome language barriers, and even enable fact checking if combined with external knowledge bases, such as SciFact and Climate-Fever.

Chart question-answering and chart understanding are recognised as distinct subtasks of VQA, leading to the development of specialised datasets, models, and applications [4, 3]. We have used geographical maps as input to some of these systems, but our preliminary tests show that existing systems do not generalise well to such input. Given the importance of maps to Earth System Sciences, and their complex nature as representational artefacts, we propose the development of an open dataset specifically designed to facilitate map QA, along with a baseline architecture to demonstrate its potential.

II. *Incubator Project description*

The aim of the project is to create an open dataset of thematic map images, accompanied by a baseline model for implementing map-based QA systems. The total expected duration is six months. Below is a description of the individual tasks that make up the project and their respective durations. Some tasks will overlap, please see Fig. 1 for reference.

The images will be sourced primarily from maps available on Our World in Data, which catalogues global maps and makes them available under a CC BY license. We will also be sampling geographic maps from papers published in journals under permissive licenses. Collecting and pre-processing (if necessary) the images is expected to last one month.

After collecting the images, we will define the types of questions the dataset will cover (two weeks). Drawing on related literature and real-world applications, we aim for the dataset to represent a broad range of potential user questions. The collected map images will then be annotated by linking maps to the corresponding questions and answers. Additionally, we will annotate map elements, such as legends, titles, north arrows, and scale bars and, where possible, the geographic extent. For this task, we will use the annotation tool LabelMe. The annotations will be provided in the form of JSON files. This phase is expected to take a total of two months. In case additional data is needed but time is constrained, data augmentation techniques such as synthetic question generation will be used to expand the dataset.

Subsequently, we will train a predictive model in an open manner, making all data and code available on Github and Hugging Face for community access. The purpose of the model is to serve as a baseline against which map QA capabilities can be demonstrated. For the model training we will use PyTorch. We will develop our baseline architecture using open-source machine learning models for both image (e.g., ResNet) and text (e.g., BERT) processing. This will take a total of seven weeks, three of which will overlap with the data annotation phase. Should the time available prove insufficient, we will fine-tune a pre-trained model.

The model's performance will be evaluated in terms of accuracy, consistency, and relevance in answering questions and in user testing with unseen input. The total duration of this phase will be three weeks, with one of these weeks overlapping with the final phase of model training, should adjustments be required.

In the final month of the project, two weeks are allocated for documenting the work, and an additional two-week buffer period is reserved for unforeseen obstacles. The documentation of the work will include the creation of a separate metadata file in an appropriate format, in accordance with the FAIR principles. If the buffer is not needed, it will be re-allocated to pursuing our stretch goal of deploying the application on Hugging Face spaces.

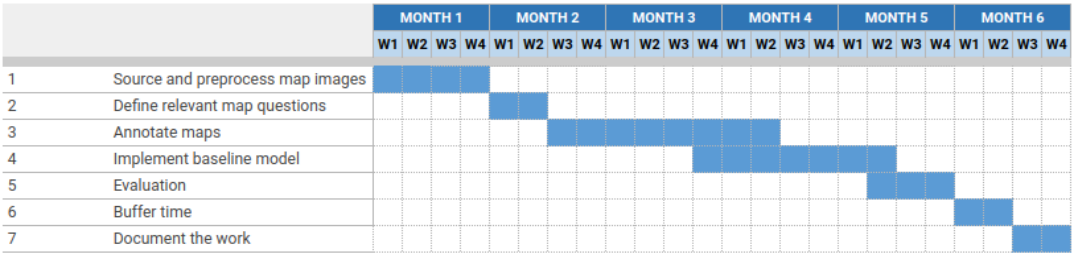

Figure 1: Gantt chart of the project.

III. *Relevance for the NFDI4Earth*

This project will deliver a first-of-its-kind dataset dedicated to map QA, together with an open source baseline model architecture, in line with the FAIR principles and against the trend of data privatization. Expected users include scientists, data curators, as well as infrastructure providers and system integrators, who can use these resources to advance map-based QA applications. For example, this dataset can support the development of tools for visually impaired users and enable the linking of geovisualisations to NFDI4Earth data repositories, thereby improving the accessibility and usability of geographic data. Maps under *Maps and Visualizations* of the Climate Service Center Germany could also serve as input to the dataset, if granted permission to use them.

IV. *Deliverables*

- Open dataset including map images, annotations and metadata.
- Open source baseline model for map QA.
- Jupyter notebook showing the usage.
- Documentation and performance evaluation report.

All items to be made publicly available on Github and Hugging Faces.

V. Finance plan

The requested funds will be for a full-time position for a period of 6 months.

References

[1] Eftychia Koukouraki and Christian Kray. A systematic approach for assessing the importance of visual differences in reproduced maps. *Cartography and Geographic Information Science*, 0(0):1–16. ISSN 1523-0406. doi: 10.1080/15230406.2024.2409920.

[2] Tung Le, Huy Tien Nguyen, and Minh Le Nguyen. Multi visual and textual embedding on visual question answering for blind people. *Neurocomputing*, 465:451–464, November 2021. ISSN 0925-2312. doi: 10.1016/j.neucom.2021.08.117.

[3] Fangyu Liu, Francesco Piccinno, Syrine Krichene, Chenxi Pang, Kenton Lee, Mandar Joshi, Yasemin Altun, Nigel Collier, and Julian Martin Eisenschlos. MatCha: Enhancing Visual Language Pretraining with Math Reasoning and Chart Derendering, May 2023. arXiv:2212.09662 [cs].

[4] Ahmed Masry, Xuan Long Do, Jia Qing Tan, Shafiq Joty, and Enamul Hoque. ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2263–2279, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-acl.177.

[5] Simon Scheider, Jim Jones, Alber Sánchez, and Carsten Keßler. Encoding and Querying Historic Map Content. In Joaquin Huerta, Sven Schade, and Carlos Granell, editors, *Connecting a Digital Europe Through Location and Place: Lecture Notes in Geoinformation and Cartograph*, June 2014. Springer, Cham. doi: 10.1007/978-3-319-03611-3_15.