Environmental Computational Science and Earth Observation Laboratory

Master thesis subject

Visual Question Answering: Assessing ground conditions after a flooding event by asking questions to images

Context

Following a large-scale natural disaster, the promptness of emergency responders is often key in providing efficient help. However, access to damage areas is often restricted if not entirely blocked in the first hours or days after the event. In this case, gathering an initial assessment of the situation at ground level could be obtained from aerial imagery, in particular drone images. With this new source of data comes a large quantity of useful information, given one has time to analyse all images and in particular possess the skills to do it efficiently. Time and skills are limiting elements that may in turn delay first aid responders on site. While artificial intelligence, especially deep learning, has been more and more proposed as a solution in assessing large quantities of images, it requires specific, specialised skills. Providing tools for a larger panel of users to use efficiently the images of damage areas rapidly would surely improve emergency response.

For a few years, a new AI framework has risen in computer vision (CV), joining elements from image and natural language processing (NLP): Visual Question Answering (VQA). Initially, VQA has been proposed as a solution to enable visually impaired people to make inquiries about their surroundings. The idea is for the user to ask a question in natural language about a visual input, images. The system needs to understand the question, analyse the image and provide an answer in natural language as well. In practice, the problem is often built as a classification task, where each possible answer is a different class. More generally, solving the VQA problem would allow a more diverse panel of users to work on and extract useful information from images, in a sense democratizing image analysis techniques.

ECEO

EPFL ENAC IIE
Devis Tuia, Prof.
Rue de l'Industrie 17
Case Postale 440
CH – 1951 Sion

Téléphone : +41 21 69 382 83 (secr).
E-mail : devis.tuia@epfl.ch

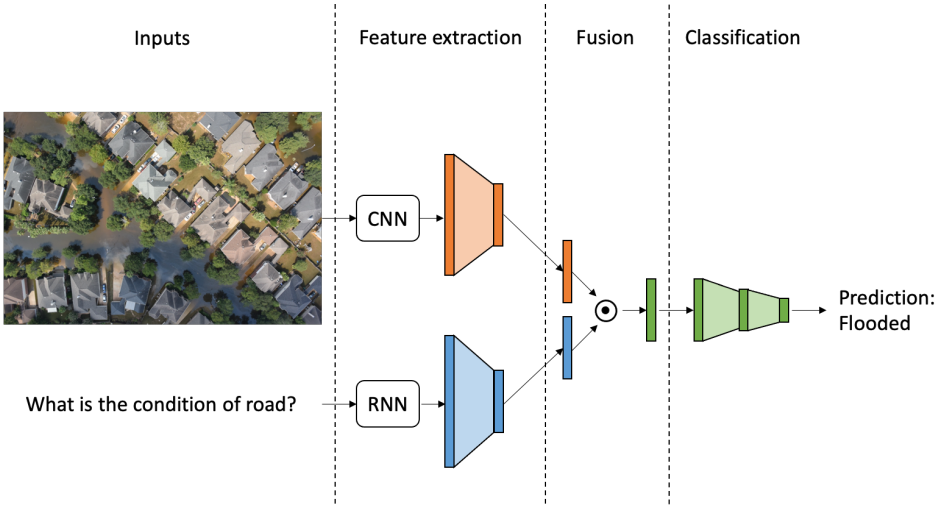

Figure: baseline architecture inspired by Lobry et al. (2020)

In the context of the flooding events, we have the chance to use a newly proposed dataset that was initially presented in the publication Rahnemoonfar et al. (2020) as well as in the FloodNet challenge for EarthVision 2021, a large-scale computer vision event for remote sensing imagery. In this project, we propose to get familiar with the task of VQA, the baseline architecture suggested by the authors of the dataset and the model of the winner(s) of the challenge. Different directions to pursue for the project can be discussed with the student, depending on interest and abilities. While improving the model in terms of performance is desirable, we especially encourage the curiosity and motivation to adapt it, leveraging the booming literature in this domain.

Objectives

- Familiarise and train a VQA architecture
- Investigate the possibilities to adapt the model and improve performance

Requirements and practical info

- Background in machine/deep learning is welcome.
- Programming skills in Python.
- The thesis will be supervised from the Sion campus.
- Access to parallel computing resources is provided.

Literature

- M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. Murphy, “FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding,” arXiv:2012.02951 [cs], Dec. 2020, Accessed: Apr. 01, 2021. [Online]. Available: http://arxiv.org/abs/2012.02951
- S. Lobry, et al. "RSVQA: Visual Question Answering for Remote Sensing Data." IEEE Transactions on Geoscience and Remote Sensing, 2020

Contact

Prof. Devis Tuia, devis.tuia@epfl.ch