AUGMENTED REMOTE SENSING WITH SMART HYBRIDIZATION FROM EXOGENOUS DATASOURCES

Louis Ulmer[⋆], *Bastien Nespoulous*[⋆], *Pierre-Marie Brunet*[†],

[⋆]Augmented Data Competence Center, Thales Services Numériques, France
[†]Centre National d'Etudes Spatiales, Toulouse, France

ABSTRACT

In the rapidly evolving field of remote sensing, the integration of diverse data sources is crucial for enhancing the accuracy and reliability of geospatial information. This paper introduces a novel approach to augment remote sensing capabilities by smartly hybridizing exogenous data sources, thereby improving the overall quality and utility of geospatial analysis. Our methodology leverages advanced Deep Learning techniques to efficiently combine heterogeneous data, such as satellite imagery, with supplementary information from external sources, including social media, news articles, and open data repositories. The application of our smart hybridization approach to natural disaster management, particularly in the context of the Turkey and Syria earthquakes from February 2023, holds immense potential for improving the effectiveness and efficiency of response efforts.

Index Terms— Data Fusion, Deep Learning, Remote Sensing, Multimodal AI, Crisis Informatics

1. INTRODUCTION

Satellite images are a valuable source of information for understanding what's happening in the field. In recent years, we have seen a significant increase in the number of services based on data from remote sensing satellites such as AI4GEO[1]. These services can be used to track environmental evolution, monitor agricultural production and assess the impact of natural disasters. Crisis informatics is a field of study that focuses on the use of information and communication technologies (ICTs) during crises and disasters. Over the years, several impactful papers have been published in this field, highlighting the importance of ICTs in disaster response and recovery efforts[2, 6, 5]. The proposed framework employs a modular, scalable architecture that facilitates seamless data fusion and harmonization, while maintaining a user-focused design for intuitive interaction and customization. By incorporating machine learning algorithms and state-of-the-art data processing tools, our approach enables the extraction of valuable insights from complex, multi-dimensional datasets, ultimately leading to more informed decision-making in various application domains, such as environmental monitoring, disaster management, and urban planning. Through rigorous testing and quality assurance measures, we demonstrate the effectiveness of our smart hybridization techniques in enhancing remote sensing capabilities and reducing uncertainties associated with data interpretation. From a use case perspective we can obtain a more comprehensive and timely understanding of the extent and severity of the earthquake impacts. The needs and vulnerability highlight the importance of continuous learning and collaboration among software engineers, domain experts, and end-users to ensure the successful implementation and adoption of this innovative approach. By embracing a problem-solving mindset, our work is a contribution for the next generation of remote sensing solutions that are more accurate, reliable, and user-centric.

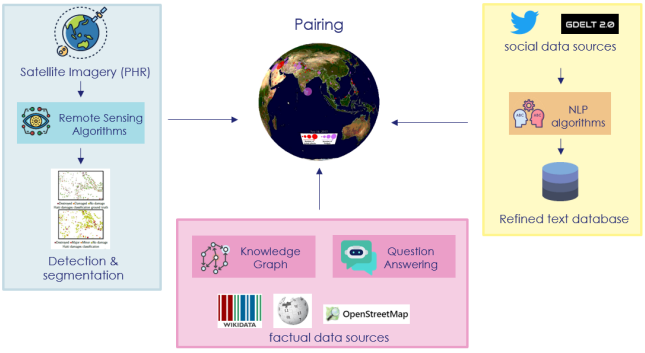

Fig. 1. Simplified diagram of our approach

2. USECASE

On February 6, 2023, earthquakes measuring 7.7 and 7.6 hit South Eastern Turkey and Syria, affecting 10 cities and resulting in more than 59,000 deaths and 121,000 injured as of February 21. We selected this example because it offers several advantages in terms of data availability: Coverage of the event by the local and international press is significant making it a good exercise to ensure the robustness of NLP methods to different languages; Information coverage via social

networks is very dense. Turkish is a well supported language with available pretrained Transformers models. This ensures that eyewitnesses testimonies can also be analysed.

2.1. Remote Sensing Data

Remote sensed data is an important component of our system, it allows to precisely locate a change in landscape after an event. We chose to use PHR images from Pleiades to obtain highly precise observations with a resolution of up to 50 cm.

2.2. Exogenous Data

Reddit is a social media platform where users can share and discuss content on various topics. It has a large online community and provides a rich source of data for researchers and analysts interested in understanding user behavior, sentiment, and trends. Users can post text, images, videos, and links, and upvote or downvote content to indicate their preferences. Reddit also has subreddits, which are dedicated communities focused on specific topics or interests, providing a more targeted source of data for researchers. Overall, Reddit is a valuable source of data for social media analysis, market research, and other applications that require insights into online user behavior and preferences.

The Global Database of Events, Language, and Tone (GDELT) is a comprehensive database of global events and *news coverage*, spanning over 200 years of history. The database includes over 300 categories of events, such as protests, conflicts, and diplomatic relations, and covers over 65 languages and 4,000 news sources from around the world. GDELT provides a valuable source of data for researchers and analysts interested in understanding global trends, patterns, and relationships across different domains, such as politics, economics, and social issues. The database is updated in real-time, allowing for near-instantaneous analysis of current events and trends.

Twitter is a popular social media platform that allows users to share short messages, or tweets, with their followers. With over 330 million active users, Like Reddit, Twitter is a wealth of data for understanding the behavior and opinions of a population. The platform features a variety of content formats, including text, images, and videos, and allows users to use hashtags and mentions to connect with other users and join conversations. Twitter also includes a system of trending topics, which are popular hashtags or keywords that are currently being discussed by users. These trending topics provide a valuable source of data for researchers interested in tracking real-time events and trends.

Wikimedia is a vast ecosystem of knowledge repositories, including Wikipedia, Wikidata, and WikiNews. Researchers can use these platforms to extract, process, and integrate relevant information to construct an event knowledge graph. Web scraping and natural language processing can extract event-related data from Wikipedia articles. Wikidata, a structured knowledge base, contains over 100 million entities and can enrich the event knowledge graph. WikiNews provides event-specific details, such as dates, locations, and participants. By leveraging these platforms, researchers can construct a comprehensive and accurate event knowledge graph.

3. PROPOSED FRAMEWORK

The proposed framework aims to enhance the information available in satellite imaging with 2 components:

- Leveraging eyewitnesses and news coverage for immediate insitu analysis; Effective rapid response to emerging crises requires near-realtime information ground. Reddit and Twitter provide authentic and instantaneous observations of the crisis. GDELT offers crucial capability to essentially harness the world's news media into a single codified information stream that brings together every piece of information that is being reported about a given crisis into a single unified data format, fully georeferenced and placed into the global context. We employ BigQuery and GDELT custom wrappers in python to gather data from various sources, including GDELT GKG, GGG, GEO, and DOC APIs.

- Creating a multimodal knowledge graph[4, 7] that provides contextual enrichment to the observed event; Handset to GDELT, we leverage the official APIs of Wikipedia, Wikidata, WikiNews, and curated Tweets and articles to collect relevant data pertaining to a given event. Specifically, we utilize these APIs to extract factual information about the event, such as its location, causality, power, and destruction.

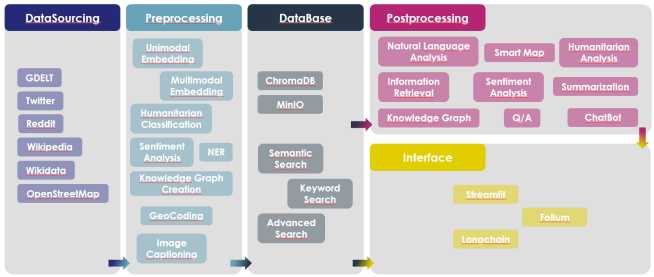

Fig. 2. An overview of our multimodal framework.

The whole pipeline is shown in figure 2. In this proposed framework, we present a comprehensive approach consisting of five distinct stages, each comprising various processes and subprocesses. The first stage, referred to as "data datasourcing," involves the extraction of diverse datasources to obtain a variety of data.

The second stage focuses on the preprocessing of multimodal data, specifically employing natural language processing (NLP) algorithms such as classification and geocoding, multimodal embedding, image captioning, and the construction and enhancement of a multimodal Knowledge Graph. This step encompasses the creation of the multimodal knowledge graph, followed by reasoning processes that merge entities and expand the graph by incorporating new subgraphs.

The third stage is dedicated to database management. In our approach, we utilized ChromaDB for embedding storage and management, while MinIO was employed for data storage. The combination of these two solutions enables the creation of a robust multimodal database.

Moving on to the fourth stage, known as post-processing, we incorporate sentiment analysis, Knowledge Graph summarization, and Question Answering techniques applied to the Knowledge Graph.

Finally, the fifth stage encompasses a comprehensive process called HMI (Human-Machine Interaction), which amalgamates the outcomes of the preceding stages. This process generates a user report and fosters interaction between the user and the system.

4. MULTIMODAL MODELING

4.1. Remote Sensed Images

We choose to deal with remote sensed images using semantic segmentation algorithms based on U-NET architectures. Remote sensing is a powerful tool for monitoring and analyzing changes in the built environment, including the detection of building collapses. In recent years, deep learning algorithms have shown great promise in improving the accuracy and efficiency of building collapse detection from remote sensed images. One such algorithm is the U-Net, a convolutional neural network architecture that has been widely used for image segmentation tasks. We decide to leverage the work done during the xView2 Challenge. The idea is to retrieve precise coordinate of interests for furter analysis and correlation with exogenous data.

4.2. Multimodal Information Retrieval

Self-supervised deep learning methods such as contrastive learning and non-contrastive learning can be used to build a joint representation between multiple modalities. Contrastive learning methods involve creating a joint representative space using two encoders of different modalities, and comparing the output vectors of positive and negative pairs using a distance-based similarity measure. Non-contrastive learning uses only positive pairs of data for learning. In this study, we have opted for an architecture such as CLIP, which enables us to reuse some of our fine-tuned models on spatial images.

4.3. Sentiment Analysis

The study's sentiment analysis component will rely solely on social media platforms, such as tweets and Reddit messages, to capture public opinion. The sentiment analysis will be limited to classifying sentiments as positive, neutral, or negative, and will use Roberta-type masking learning models. These models will utilize pre-trained models available from the open source community. The study posits that this model type's performance is more than sufficient for analyzing the impact of an event on the population.

4.4. Image Preprocessing

To extract entities from the image modality, we can select part of images as entity using detection models or segmentation models or transform image into text to extract textual information from the image. We have decided to preprocess images using BLIP2, taking advantage of the fact that we already have finetuned weights of this model on spatial images.

4.5. Knowledge Graph Creation

In this paper, we focus on the automated methods to extract Knowledge Graph[4]. We investigate the performance of different NLP techniques for social network analysis, Wikipedia, and user generated content. For social network analysis, we observe that dependency rule methods are not effective due to the poor detection of sentence syntax. Instead, we employ named entity recognition (NER) methods, which are better suited for processing tweets and perform well due to the factuality of the tweets approaching relevant information. Regarding Wikipedia, we find it easier to construct a dataset of sentence triple pairs, and therefore, we use model-based methods, such as Bart and T5. Finally, for user-generated content and media, where the user can fill in to complete the knowledge graph, we use dependency-based methods to avoid depending on Wikipedia's writing style via language models methods, which may not be able to extract complex triplets from long texts.

4.6. Knowledge Graph Post-Processing

The main objective of this project is to explore the linearization of knowledge graphs, focusing on the simplicity of its implementation. Initially, we have chosen this approach over other methods because of its ease of use and direct implementation. In future work, we plan to investigate the use of embedding to transform knowledge graphs into text using Graph Neural Networks embedders[7]. Theirs approaches will be applied on generative language models like BART or T5 for summarization and question answering on a Knowledge Graph.

URGENT AID TO HATAY LOC ! MY GRANDMOTHER AND MY BROTHER-IN-LAW ARE STILL UNDER THE RUBBLE NO ELECTRICITY, NO AFAD, IT'S FREEZING, PLEASE HEAR OUR VOICES ! General Şükrü Kanatlı Mahallesi LOC 12th Street LOC , Bereket Apartmani LOC Kat: 3 apartment: 5 ANTALYA LOC / HATAY LOC @hatay

Fig. 3. NER algorithm applied to a tweet that has been translated to english, we can see that the address has been successfully detected. Here we used GPT3.5.

5. GEOREFERENCING

Since Remote Sensed Data is fully georeferenced, being able to georeference exogenous data is a critical part of our project. Specifically, georeferencing techniques for social media platforms like Twitter and Reddit, as well as press articles allow spatial analysis and visualization alongside building wreckage segmentation algorithms. For Twitter, geotagged tweets containing latitude and longitude coordinates can be directly plotted on a map. However, many tweets lack this information. Similarly, Reddit posts and comments can be georeferenced by leveraging subreddit-specific geolocation but the location accuracy remains poor. Press articles can be georeferenced thanks to GDELT GGG API that is already preprocessed with established NLP and rule based algorithms.

Named Entity Recognition (NER) is a subtask of Natural Processing (NLP) that involves identifying and classifying named entities in text, such as people, organizations, and locations. By using NER, we can extract location names from social media posts and press articles, and then use geocoding techniques such as [3] to convert those location names into latitude and longitude coordinates. This allows us to perform spatial analysis and visualization on a wide range of data sources, even if they do not contain explicit geolocation information. We also investigate keyplaces detection algorithms on images extracted from press and social media. Once geolocated, these text data points can be plotted on the same map, enabling ways to study spatial patterns in communication, information dissemination, and public sentiment across various plateforms. This approach has numerous applications, including disaster response, public health monitoring, and understanding the geographic distribution of opinions and trends.

6. EXPECTED RESULTS

As the study is still ongoing, we won't show results here but more discuss about our expectations to be reached next november 2023. Our expectations are that we will seamlessly blend results from segmentation on remote sensed images while providing a comprehensive analysis of the exogenous data.

When looking at 4, we can get a clear indication of the areas that require support. Moreover we expect to grow our understanding of what precisely happened by conducting topic modeling, sentiment analysis, automatic summarization, Q/A on our large source of data.

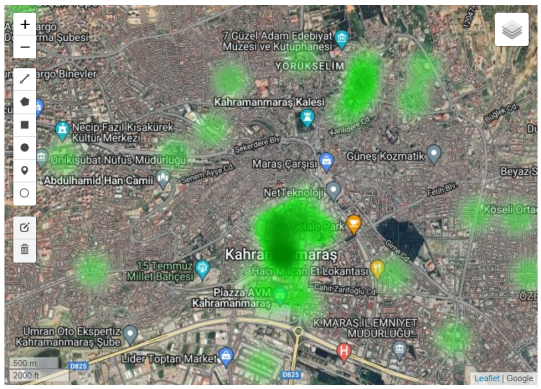

Fig. 4. Density map of messages on social networks containing information about injured or dead people around kahramanmaraş city

REFERENCES

[1] P.M. Brunet, S. Baillarin, P. Lassalle, F. Weissgerber, B. Vallet, C. Triquet, G. Foulon, G. Romeyer, G. Souille, L. Gabet, C. Ferrero, T.-L. Huynh, and E. Lavergne. Ai4geo: A path from 3d model to digital twin. In *2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022.

[2] C. Fan, C. Zhang, A. Yahja, and A. Mostafavi. Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. *International Journal of Information Management*, 2021.

[3] S. Hoffmann. Nominatim : Open-source geocoding with openstreetmap data, 2020.

[4] M. Li, R. Xu, S. Wang, L. Zhou, X. Lin, C. Zhu, M. Zeng, H. Ji, and S. Chang. Clip-event: Connecting text and images with event structures. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[5] Q. Ning, B. Zhou, H. Wu, H. Peng, C. Fan, and M. Gardner. A meta-framework for spatiotemporal quantity extraction from text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022.

[6] K. Saldana Ochoa and T. Comes. A machine learning approach for rapid disaster response based on multi-modal data. the case of housing & shelter needs. 2021.

[7] Y. Zhou and G. Long. Multimodal event transformer for image-guided story ending generation. In *Association for Computational Linguistics*, 2023.