Empowering Crisis Response Efforts: A Novel Approach to Geolocating Social Media Images for Enhanced Situational Awareness

Hafiz Budi Firmansyah*
Centre Universitaire d'Informatique
University of Geneva, Switzerland
Hafiz.Firmansyah@etu.unige.ch

Jose Luis Fernandez-Marquez
Centre Universitaire d'Informatique
University of Geneva, Switzerland
JoseLuis.Fernandez@unige.ch

Mehmet Oguz Mulayim
Artificial Intelligence Research Institute, IIIA
Spanish National Research Council, CSIC
Cerdanyola del Vallès, Spain

Jorge Gomes
Virtual Operations Support Teams
VOST Europe
Santarém, Portugal

Joao Ribeiro
Virtual Operations Support Teams
VOST Europe
Santarém, Portugal

Valerio Lorini
Joint Research Center
European Commission
Ispra, Italy

ABSTRACT

Social media and online platforms play an important role in assessing the impact of a natural disaster, especially in the immediate aftermath of the event. The information from social media demonstrated the potential to accelerate the response to a crisis. Locating social media images is critically important to help the victims immediately, mobilize community support, and provide enhanced situational awareness. However, finding the location of relevant social media images still challenges humans and computational processes.

This study introduces a social media image classifier aimed at enhancing crowdsourced geolocation. The model is trained using data annotated by experts in disaster risk management. Its main goal is to reduce the difficulty of geolocating images by detecting those that are "easy" to geolocate. We combined the classifier with an ad-hoc crowdsourcing platform and tested it with pictures posted during a crisis. The experimental results indicate that the proposed approach speeds up the geolocation process of social media images while increasing the level of location precision

Keywords

Social media, geolocation, disaster management

INTRODUCTION

Data from social media demonstrates the ability to facilitate quicker response in crisis situations. In 2020, the Joint Research Center European Commission activated a social media analysis to understand the Beirut explosion. Their work proposes a framework to strengthen the use of social media for crisis management that could complement the current Copernicus Mapping service in the initial phases of the crisis. Further developments aim at the automated

*corresponding author

detection of relevant media, their localization and classification providing an initial impact assessment starting from a few moments after the disaster [1]. Despite usefullness of social media data in previous example, employing data from social media to evaluate the damage introduces significant challenges. For privacy concerns, major social media platforms remove the meta information about locations by default. Out of millions of posts on Twitter, one of the most used platforms by researchers, only up to 3% are geolocated by users (Huang and Carley, 2020). Furthermore, identifying the location of relevant photos with evidence of damage to organize the first response efficiently presents a challenge. While several techniques have been proposed to automatically geolocate social media content, including images (e.g., Murgese et al., 2022) and text (e.g., Scalia et al., 2022), their predictions need to be more precise to enable their use by practitioners in the field. Combining automatic geolocation with crowdsourcing offers a valuable solution to this problem, with the former providing a preliminary location estimate and the latter being used to enhance the level of precision of that location. The geolocation is then pinpointed on a map to visualize the affected areas. The actual goal of extracting location from social media images is to create a disaster map that could help the first responders locate the events seen on posted images.

Thanks to the combination of Crowdsourcing and Artificial Intelligence methods, this research aims to reduce the overall time of geolocating social media images posted from regions impacted by a crisis and increase the geolocation precision. To pursue the objective, we decrease the difficulty of geolocating images by selecting easy-to-geolocate images. The study leverages past research activities on social media images for the disaster risk management domain to support emergency response as a framework (Rufolo et al., 2021), a platform (Lorini, Panizio, and Castillo, 2022), and a disaster map (Lorini, Rufolo, and Castillo, 2022).

The remainder of the paper is organized as follows: First, we discuss related work and research gaps. Second, we describe our dataset and experiment method based on the research gaps and present the methods we followed to shorten the research gaps and detail the experiments that demonstrate the validity of our approach. Then, we report the experimental results with social media data from four disaster events. Finally, last section concludes this work and provides future research directions.

RELATED WORK

Geolocation prediction on social media involves linking an entity, like images or posts, to a specific location with a high level of accuracy. This process can pinpoint locations ranging from a street level to city or country-wide areas. In recent years, the scientific community has shown a growing interest in using social media geolocation for disaster management research. Various studies have adopted methods using images, text, and combined approaches to improve the accuracy of geolocation identification.

In an image-based approach,(Murgese et al., 2022) proposed a solution to estimate images' locations by implementing Focal Modulation Networks for predicting the geolocation. The authors used geolocated images from Flickr[2] and Mapillary[3] to evaluate the proposed approach. Mapillary is an open-source map that provides street-level scenes. The focus of the study was the Barcelona City, Spain. Around 70% of Mapillary data, which had street-level imagery similar to Google Street View, was successfully pinpointed at neighborhood level. However, the proposed approach did not use social media data, and the precision of geolocation prediction was limited to the neighborhood level.

A multimodal (combining visual and textual features) approach was introduced by (Tahmasebzadeh et al., 2023). In the work, the author predicted the geolocation of news photos. To implement the approach, the authors also introduced a new dataset called Multimodal Geolocation Estimation of News Photos (MMG-NewsPhoto). The work demonstrated that the multimodal approach outperformed the unimodal.

Related works on text-based geolocation use several approaches, for instance, Named Entity Recognition and Location Indicative Words, to infer the geolocation candidate (Utomo et al., 2018). For example,(Middleton et al., 2018), implemented the Named Entity Recognition algorithm to predict geolocation from social media data. In the experiment, they used Twitter[4] and Flickr as the experimental dataset. The works were evaluated qualitatively and quantitatively. However, the experimental result demonstrated low results with a high radius from the location.

Another technique called Location Indicative Words, adopted by (Han et al., 2012), used a dataset of cities in the experiment. They developed and evaluated the approach on the English dataset. The experimental result demonstrated that the feature selection improved information gain and outperformed state-of-the-art geolocation prediction methods by 10.6% in terms of accuracy. Also, the result showed the mean and median of prediction error distance by 45 km and 209 km. Despite the innovative technique, the precision deviation was still high.

[1]https://publications.jrc.ec.europa.eu/repository/handle/JRC124081
[2]https://www.flickr.com/
[3]https://www.mapillary.com
[4]https://twitter.com

In a recent study,(Scalia et al., 2022) proposed a context-aware approach to geolocate emergency-related social media posts. The system implemented post metadata and post network of relationships. The experimental results showed that around 35% of images geolocation inferences were within a 500-meter radius from the ground-truth location. About 62% of images were predicted within 12 km. However, the results still had a high radius from the actual location.

Text-based geolocation prediction research often includes text from social media posts. The other alternative is by extracting the text from an image and using the extracted text to retrieve the location. (Firmansyah et al., 2023) proposed an automated text-extraction pipeline to predict the geolocation of an image. The work used the text in a social media image as input for a text-based geolocation prediction algorithm. The experiment incorporated 1,752 images curated from four different disasters in the world (Croatia Earthquake, Catania Floods, Central Europe Floods, and Haiti Earthquake). The experimental results demonstrated that the number of original images that could be successfully geolocated within the bounding boxes (affected area) varies between 9% and 11% of the original dataset, depending on the geolocation algorithm used. However, the work was still in the preliminary phases and had a country-level precision.

In a hybrid technique, (Ravi Shankar et al., 2019) introduced a platform called Crowd4EMS. This platform collects, analyzes, and geolocates social media information by combining crowdsourcing and automatic methods. They used data related to the Amatrice Earthquake in 2016 to evaluate the platform, coming from Flickr, Twitter, and Youtube. The work demonstrated that the suggested approach could better support the crowdsourcing communities in providing high-precision geolocation in the context of disaster response. However, the work also shows that crowdsourcing the geolocation of social media images is a complex and time-consuming task, which produces unpredictable delays that could prevent the information from being available in the immediate aftermath of the event.

In addressing the challenge of combining expert analytical capabilities with the scalability of crowdsourced efforts, the study introduces "GroundTruth," an online system designed to augment expert tasks with crowd participation. This system employs three distinct shared representations—a diagram, grid, and heatmap—to facilitate real-time collaboration between experts and crowds in the complex task of image geolocation, a critical process utilized by professional journalists and human rights investigators. Through a mixed-methods evaluation involving 11 experts and 567 crowd workers, the findings demonstrate GroundTruth's effectiveness in aiding image geolocation tasks, while also highlighting the interaction challenges and successful strategies between experts and non-experts. The research contributes valuable insights into the design of shared representations for enhancing visual search and sensemaking tasks, extending its implications beyond image geolocation to broader applications (Venkatagiri et al., 2019).

RESEARCH QUESTIONS

This paper proposes that the timing and accuracy of the overall geolocation process of social media posts using crowdsourcing and automatic methods can be improved by prioritizing those images deemed easy to geolocate. In this way, easy-to-geolocate images are presented earlier to the crowd, preventing delays from difficult or even impossible-to-geolocate social media posts. This process requires the implementation of an image classifier to detect the level of difficulty associated with the process of geolocating an image.

This work answers the following research questions (RQs):

- **Can the proposed classifier automatically predict the difficulty of geolocating an image?** (RQ1)
- **Does favoring images easy-to-classify improve the overall geolocation process?** (RQ2)

DATASET, METHODOLOGY, AND EXPERIMENT

This section provides an overview of the dataset and methods utilized in the experiment, followed by a comprehensive explanation of the experimental design. To answer the proposed research questions, we devised three major tasks: In the first task, images extracted from a reference dataset are presented to a crowd of domain experts that labels each image with a grade representing the difficulty level of geolocalization. To answer the RQ1 stated above, the second task aims at training a deep learning model for classification using the data annotated in the first task. Finally, the third task consists of a domain adaptation experiment where the same crowd of experts is asked to geolocate previously unseen images annotated by the model, thus answering the second research question (RQ2).

Table 1. Summary of the image geolocating results considering just the total number of geolocatable records

Records	Catania floods		EU floods		Haiti earthquake		Croatia earthquake		All datasets	
	N	%	N	%	N	%	N	%	N	%
Geolocatable	113	100	271	100	1,462	100	627	100	2,473	100
Geolocated	70	61.9	214	79	961	65.7	549	87.6	1,794	72.5
Timeout	43	38.1	57	21	501	34.3	78	12.4	679	27.5

Dataset

As a reference dataset for this research, we used the images geolocated by GIS (Geographic Information System) experts for the study EMSV070 - Reference Datasets for SM (Social Media) Image Geolocation in DRM (Disaster Risk Management) requested within the Copernicus Emergency Management Services framework. The dataset contains geolocated images and text collected during four different disasters: (1) Croatia Earthquake, 2020 (30.7%); (2) EU Floods, 2021 (3.9%); (3) Central European floods, 2021 (11.9%); and (4) Haiti Earthquake, 2010 (53.5%) (see Table 1).

Preferably, the damage in the image (flooded point, damaged building, etc.) was selected for geolocation. There are 5,430 records in the four datasets together. Of these, almost 13% (693) were repeated (non-analyzed), and about 87% (4,737) were analyzed. The details of 4,737 records were 2,264 non-geolocatable and 2,473 geolocatable. Out of 2,473 geolocatable records, we opted for 1,794 records, while 679 records were excluded because they timed out during image access.

Within the analyzed records, the non-geolocatable images were identified and discarded by GIS experts. Reasons for being considered as non-geolocatable are:

1. The URL of the image does not work, or the URLs of both the image and tweet do not work.
2. The information visible on the image does not allow its geolocation since it is a mosaic of images or a painting, or the image has bad quality (e.g., it is blurry, dark, etc.), or the image does not have visible, recognizable elements to be geolocated unless external information helps in its geolocation.
3. The image could be geolocated only at the city level.
4. The image does not correspond to any of the disaster events under study.
5. The image corresponds to regions outside of the study areas.

The geolocation was performed by GIS experts considering the following:

1. The information provided by the images themselves.
2. The text contained in the tweets.
3. The images and/or videos contained in the tweets.

Google Street View, Mapillary, Google Earth Pro, and Google Maps were used for localizing the places, recognizing the visible elements in the images, and/or for the subsequent image geolocation. Additionally, the already recognized images during the execution of the task were also considered for the geolocation of new images. Mainly, the geolocation was performed by:

1. Focusing, when possible, on the geographic information specified in the text of the tweets.
2. Selecting in the image some location relevant to the disaster event under study (flooded area or damage due to an earthquake), which was expected to be easily geolocated with the help of Google Street View or Mapillary, etc.
3. Selecting some feature/s in the image that help locate the photo the viewers used.
4. Providing the [Latitude] and [Longitude] for the selected location.

Consequently, in our work, we used images that were geolocated by the experts at least at the level of city and in 90% of cases with a precision of 10 to 100 meters. Figure 1 depicts the geolocation accuracy distribution of four different disasters.

Accuracy (m)	Catania Floods		EU Floods		Haiti Earthquake		Croatia Earthquake		All datasets	
	N	%	N	%	N	%	N	%	N	%
<10	65	92.9	173	80.8	647	67.3	532	96.9	1,417	79
<100	1	1.4	25	11.7	213	22.2	17	3.1	256	14.3
<1000	4	5.7	16	7.5	101	10.5	0	0	121	6.7
>1000	0	0	0	0	0	0	0	0	0	0
Total	70	100	214	100	961	100	549	100	1,794	100

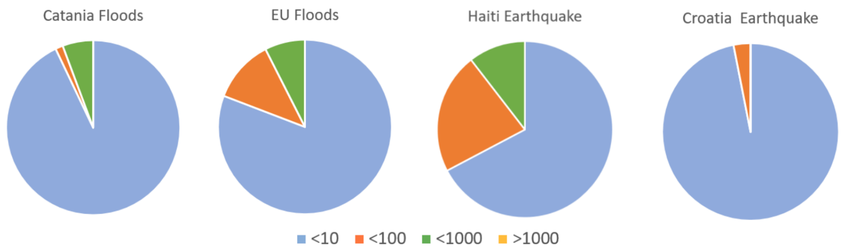

Figure 1. Distribution of the accuracy values(%) assigned to the images geolocated in each of the four datasets

We asked experts in the domain of Disaster Response, specifically the European network VOST Portugal (VOSTPT) to classify the images of the reference dataset, each based on the difficulty of geolocating them. VOSTPT is a community of citizens from different countries accross Europe that has an objective to create a more prepared, more informed, and more resilient society to natural disasters and other events [5]. To label the images, four different levels were used: (1) Easy, (2) Medium, (3) Difficult, or (4) Impossible to geolocate (see Figure 2). Each image is labeled at least by five different raters.

Figure 2. The example of geolocation difficulty level for image

The crowdsourcing of the images was carried out via the Pybossa framework (González et al., 2020). Specifically, the execution of this task uses an instance of Pybossa hosted at the Citizen Science Center Zurich, a joint initiative created by the University of Zurich and ETH Zurich[6]. The application is called the Citizen Science Project Builder (CSPB).

[5]https://vosteurope.org/
[6]https://citizenscience.ch/en/

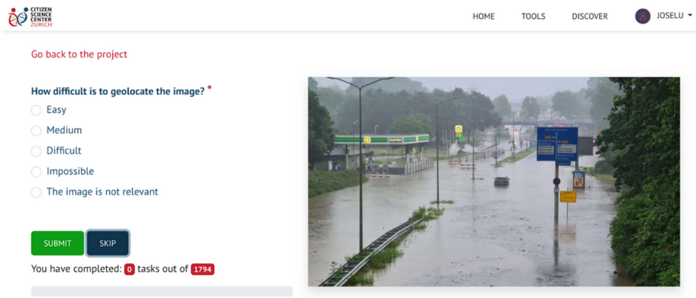

Figure 3. Screenshot of the Citizen Science Project Builder used to crowdsource the labeling of images depending on the level of difficulty to find its location

Figure 3 shows the user interface implemented for crowdsourcing the geolocation difficulty level of each picture in the dataset. The crowd is asked to determine the level of difficulty to geolocate the image on the right. The user interface also enables users to answer two different conditions stating whether it is impossible to geolocate or the image is not relevant. If the users were not sure about their choice, they could also skip the task.

As for the dataset for the second task, i.e., training our deep learning model, Majority Voting and Dawid-Skene (Dawid & Skene, 1979) consensus mechanisms were adopted to aggregate the annotations collected in the first task. We detail both mechanisms in the experimental results section.

Training a CNN to Predict the Level of Difficulty in Geolocating a Given Photo

Our second task is to create a classifier to classify the difficulty level to geolocate a given photo automatically. To this end, we implemented a Convolutional Neural Network (CNN) and the methodology for training the CNN consisted of the following steps:

First, we imported images from the annotated training dataset. We loaded images and distributed them among the classes "Easy" and "Difficult". In the beginning, we decided to use all the labels existing in the annotations, but after some preliminary analysis, we decided to reduce the complexity of the classification. Hence, we removed the Medium samples from the dataset and merged Difficult and Impossible into Difficult to make the task a binary classification between Easy (Easy) and Difficult (Difficult or Impossible). The reason for this choice was to focus on accurately identifying True Positive values (the Easy class).

We then deduplicated the data using hashing algorithms[7] that are particularly good at finding exact duplicates, as well as CNNs, which are also good at finding near duplicates. We then used some standard data analysis to check the distribution of images. Figure 4 shows the distribution of the images after the deduplication. The classes were well-balanced.

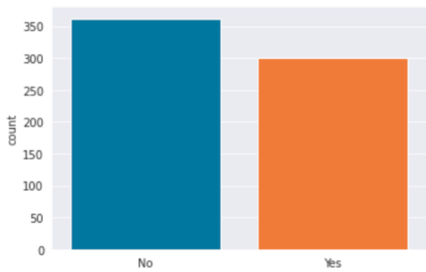

Figure 4. The almost-balanced distribution of images after deduplication

We split the dataset with a training, validation, and test ratio of 70% (448 images), 20% (150), and 10% (61), respectively. The data preparation phase included data augmentation and pre-processing. The data augmentation increased the diversity of our training set by applying random (but realistic) image rotation and image flipping (horizontally). Figure 5 shows an example of data augmentation.

[7]https://pypi.org/project/imagededup/

Figure 5. Image augmentation examples. The augmentation flips and rotates the images at certain degrees.

This step basically multiplies the number of images used in our training. Every image was augmented by being flipped and rotated. After freezing the all layers, the first 25 epochs to train the neural network. The next 25 epochs were used to fine-tune the neural network. The number of epochs was inspired by a similar experiment on leveraging a pre-trained model using social media images for natural disaster (Firmansyah et al., 2022; Ofli et al., 2020) and considering several factors such as dataset size, overfitting prevention, and learning rate.

EfficientNet [8] is among the most accomplished models that reach state-of-the-art accuracy on both ImageNet [9] and common image classification transfer learning tasks. The model was performing better than other models and could be chosen as the best option (Alam et al., 2020). In our experiment, we chose the EfficientNet-B6 due to its good balance between computational demand and model performance and because it uses an image size (528x528 resolution) similar to the ones found on social networks. Aware of the fact that training a deep learning model from scratch with a relatively small dataset might lead to suboptimal results, we implemented transfer learning. It involved starting with a pre-existing model trained on a large dataset and adapting it to our task. This practice reduces the data needed, reduces training time, and improves the model's performance on smaller datasets.

To customize the pre-trained model for our specific task, we performed fine-tuning on the last 66 layers of the EfficientNet-B6 in a two-steps process: first freezing all layers and training only the top layers, and then unfreezing the last 66 layers (out of 666). A relatively large learning rate (1e-2) was used for the first step, while the second step fit the model using a smaller learning rate. Fine-tuning involves adapting the weights of a pre-trained model to new data, allowing us to leverage both the general feature extraction learned during pre-training and the specific patterns found in our disaster imagery.

In conclusion, the experiment sought to leverage the power of EfficientNet-B6 and the robustness of noisy student pre-training by fine-tuning the model to our specific classification. Noisy student is a semi-supervised learning method where a model trained on labeled data (teacher model) produces pseudo-labels for unlabeled data. The pseudo-labels are then combined with noisy data and labeled data to train another model called the student model (Xie et al., 2020). Figure 6 shows a comparison of accuracy while testing different weights. It compared two EfficientNet-B6 models with two different weight (ImageNet and Noisy Student). The EfficientNet-B6 model used a training dataset of 448 images as explained above.

The base classifiers were trained with the objective of decreasing the cross-entropy loss. Cross-entropy measures how well a classifier approximates the probabilities of its predictions. The code generated to create and evaluate the classifier is accessible as open-source[10].

We ran two experiments with two datasets originating from Majority Voting (MV) and Dawid-Skene (DS) consensus algorithms and compared the performances of the two models. The model trained with the DS dataset presented better results than the one with Majority Voting, as we expected since DS weighs the raters' answers based on their error rates computed by the same algorithm. That is, the annotations of the raters who perform better (i.e., make fewer errors) have more influence on the consensus calculation. Thus, we expect to have a more reliable consensus with DS. Figure 7 demonstrates the comparison of the training accuracies with MV and DS datasets and their evolution after being fine-tuned (the fine tuning started at the vertical green line).

[8] https://paperswithcode.com/method/efficientnet
[9] https://paperswithcode.com/dataset/imagenet
[10] https://github.com/hafizbudi/Geolocation_AI_Crowdsourcing

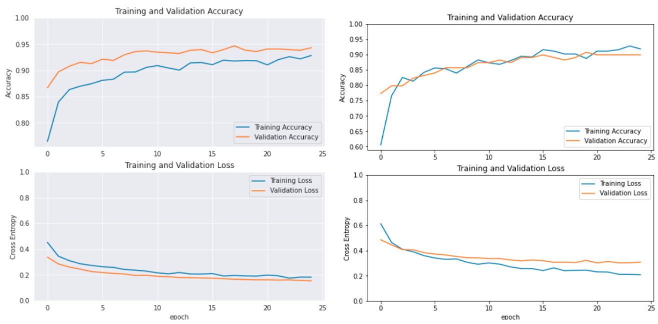

Figure 6. ImageNet (left) and Noisy student (right) weight performance comparison on model training

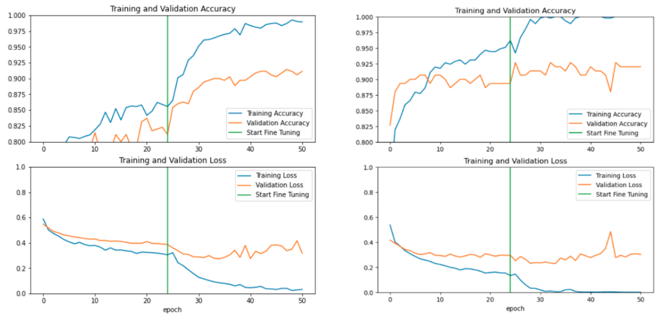

Figure 7. The comparison of model training performances between Majority Voting (left) and Dawid-Skene (right) consensus training datasets.

This preliminary result shows that classifiers could estimate the difficulty of geolocating a social media image. This step is critical to filter out images that are impossible to geolocate and to optimize the ratio of images geolocated per person right after a disaster happens when the time is crucial. Thanks to this methodology, easy images could be scheduled with higher priority than those difficult to geolocate, saving invaluable time for disaster management. We assess the validity of our hypothesis in the following section. Figure 8 shows examples of classifications correctly inferred by the model.

Domain Adaptation Classification Experiment

A central hypothesis in this research is that the photos which are automatically classified as Easy to geolocate should be geolocated with higher precision and faster by humans than the images classified as Difficult. The tool selected for the crowd annotation experiment is Citizen Science Project Builder [11] adopting the user interface design implemented within the E2MC EU project that facilitates manual geolocation (Ravi Shankar et al., 2019).

To answer our RQ2, we set up a domain adaptation classification experiment (Gopalan et al., 2015) where the model described in the previous section classifies unseen photos and human experts geolocate them and provide the precision of the location. Figure 9 and Figure 10 show screenshots from the Citizen Science Project Builder tool with the user interface for geolocating social media images. The interface combines the following: (1) initial estimation of the position of the image extracted from the analysis of the associated text to the photo, which allows participants to start refining the position rather than geolocating an image from scratch; (2) incorporation of Google Map and Google Street View which allow matching the image from social media, with images from Google Street View. When the user locates the photo on Google Street View and clicks the Submit button, the geolocation (longitude and latitude) is submitted automatically within a precision of a few meters. The user interface allows to annotate the precision of the geolocation with three different levels, i.e. the precision the user was able to achieve in the geolocation process.

Figure 11 shows the three levels of precision: i) High precision means the user found the exact location by matching it with the referenced dataset's latitude and longitude. ii) Medium precision means the user knows the image was taken close to a position, such as a street, road, path, or similar, but it could not be matched the referenced one. iii) Low precision means the image is geolocated only at the city level.

[11] https://lab.citizenscience.ch/en/tools/projectbuilder

Figure 8. Examples correctly classified by the model. The text in the upper of each image corresponds to image geolocation difficulty level

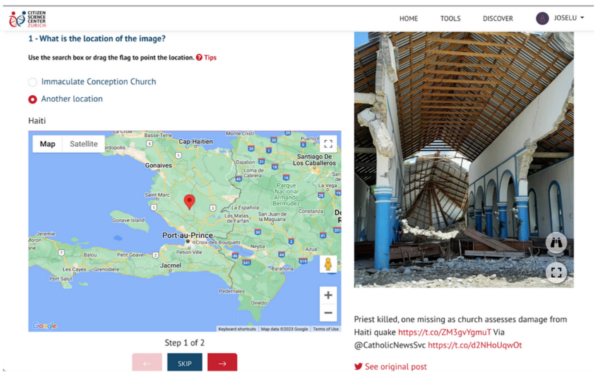

Figure 9. Screenshot of the Citizen Science Project Builder used to crowdsource geolocation of the photos (Photo from Haiti earthquake).

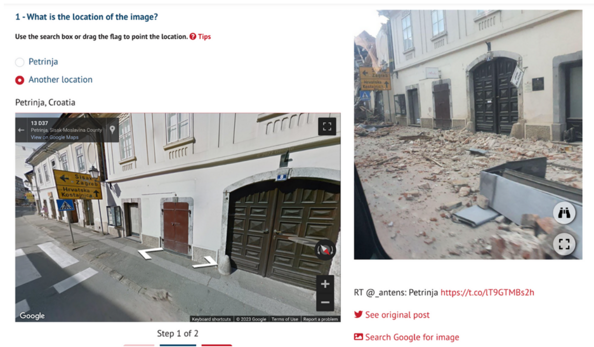

Figure 10. Screenshot geolocation using Google Street View (Photo from Croatia earthquake)

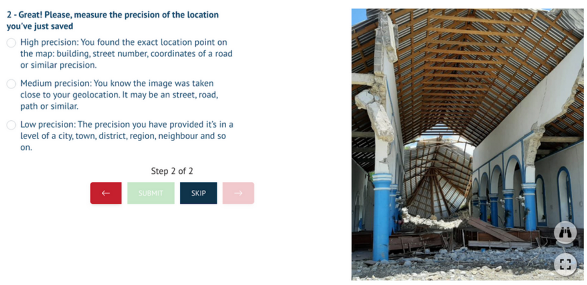

Figure 11. Screenshot CS Project Builder, final step geolocation precision levels.

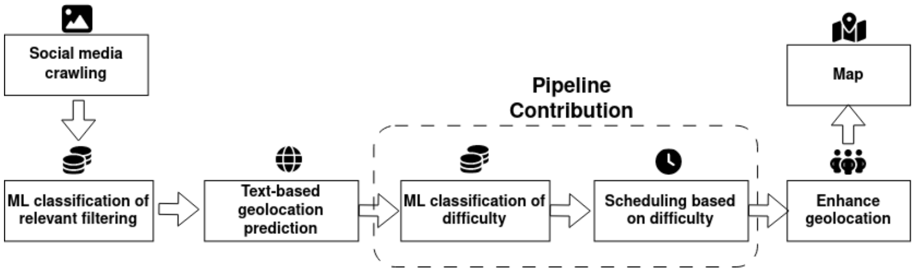

Figure 12. The pipeline of social media images geolocation

Proposed Pipeline

The most common approach to get the geolocation is by extracting location information tagged to the social media post using GPS. Our approach is completely different from that context for two reasons. First, our approach minimized the crowd effort by automatically estimating the geolocation difficulty using machine learning. Second, the approach incorporated crowd to accelerate and improve the image geolocation step.

To make use of the geolocation enhancement for the social media data proposed in this paper, we provide a novel pipeline which extends previous work from Crowd4EMS (Ravi Shankar et al., 2019) and SMDRM (Lorini, Panizio, & Castillo, 2022) by adding two additional processes. Figure 12 shows the overview of a proposed pipeline which is composed of six processes. The dashed line indicates the two additional processes integrated into the pipeline:

- **Social media crawling** fetches informative posts from social media by selecting appropriate keywords which depend on the type of disaster. The initial step is to get the social media post by formulating the appropriate search query (Alam et al., 2018; Lorini, Panizio, & Castillo, 2022).
- **ML classification of relevant filtering process** filters the irrelevant images and keeps the relevant images crawled from the previous process. For example, images containing cartoons, banners, news logos, and edited images are considered irrelevant (Alam et al., 2018).
- **Text-based geolocation prediction** uses social media posts to predict the initial geolocation of images. The main goal of the process is to provide initial geolocation information based on social media posts. The text-based geolocation prediction is used to prevent the crowds to find an image location from scratch, when possible.
- **ML classification of difficulty** predicts the difficulty level to geolocate an image into predetermined categories (i.e. easy, medium, difficult, etc). The process is critical because the institution's resources could put attention to the easier-to-geolocate and more relevant images first instead of to all images. For example, images that are impossible to geolocate should not be sent to the crowd. The process of classification difficulty is done by implementing automated machine learning (ML) described previously in Section Training a CNN to predict level of difficulty in geolocating a photo.
- **Scheduling based on difficulty** chooses images in a desired order of difficulty level to be sent to the crowd. In certain situations like disasters, the time constraint may be more strict. When time is considered an essential dimension for geolocating an image, the first responders organization should prioritize images that need to be sent to the crowd earlier (e.g., Easy before Difficult images as labeled by the automated machine learning).
- **Enhance geolocation** incorporates crowd to improve the geolocation precision.
- **Map** visualizes the geolocation enhancement on a map.

EXPERIMENTAL RESULTS

This section provides the experimental results from the three different tasks presented in Section Dataset, methodology and experiment. The aim of this section is to answer the two research questions.

As the pilot study, we first worked with 1,794 photos annotated by the VOSTPT community using the interface presented in Figure 3, with the redundancy of at least five annotations per image. In total, 8,970 classifications were performed by the VOSTPT community were each image was classified by the level of geolocation difficulty, i.e. Easy, Medium, Difficult and Impossible. With the annotation dataset in hand, we initially evaluated the inter-rater agreement. Fleiss' Kappa[12] is a common statistical measure to assess the reliability of agreement between raters in the context of categorical ratings, like our case. We computed the Fleiss' Kappa for the VOSTPT dataset using the Python library Crowdnalysis (Cerquides & Mülâyim, 2022). The value was 0.288 which indicates a Fair agreement on the kappa scale among the VOSTPT community for this dataset. The low value of fleiss kappa suggests that the task of assigning a geolocation difficulty level to a social media image is a complex task and the annotation result might be noisy. Also, there was a probability that the annotator would answer the level of difficulty randomly. Considering that issue in the experiment, we decided to choose Dawid-Skene (DS) model instead of majority voting (MV) since DS also models individual annotator behavior. We further explain this choice below.

To generate the dataset for the next step, i.e., the training of the CNN, we needed to find a consensus between the five annotators for each image. The Crowdnalysis library, besides the standard Majority Voting, provides advanced probabilistic methods of consensus (Cerquides et al., 2021) such as the seminal Dawid-Skene model (Dawid & Skene, 1979) (DS). The DS method allows modelling individual annotator behaviour, thus with enough data, it yields more reliable consensus results which can prove crucial in disaster management scenarios, as mentioned earlier.

Table 2. The percentage of warnings in consensuses with Dawid-Skene (DS) and Majority Voting (MV) models.

	Easy	**Medium**	**Difficult**	**Impossible**	**Not Relevant**
DS	**4.4%**	**5.2%**	**9.7%**	**0.7%**	0.43%
MV	57.2%	58.8%	10%	2.7%	**0%**

Table 2 shows the percentage of warnings for each consensus result. The warning was a condition where there was only a < 0.1 difference between the probabilities of the top and second-best estimated consensus classes for the photo. The boldface values indicate the lower values between models trained with DS and MV consensus datasets.

Table 3 shows the confusion matrix result of evaluating the model trained with the DS dataset. The model exhibits 0.87 accuracy, 0.88 precision, and 0.88 recall. Assessing the difficulty of geolocating an image is challenging even for humans, as demonstrated by the Kappa value above. However, the evaluation of the proposed model shows that it is possible to automatically predict the difficulty of geolocating an image. This result answered the first Research Question (RQ1).

Table 3. Confusion Matrix of the CNN model with binary classification

	Predicted	
Actual	Difficult	Easy
Difficult	29	4
Easy	4	24

After demonstrating the feasibility of predicting the difficulty of geolocating an image from social media, the next question to address is whether images classified as easy contribute to the enhancement of the overall geolocation process. The main focus is to investigate whether images labeled as easy result in increased precision and reduced time for geolocation during the crowdsourcing process. Figure 13 shows that most of the images predicted as easy are gelocated with high precision (horizontal axis = 2) and within 300 seconds. On the other hand, Figure 14 shows that images predicted as difficult are geolocated with low precision (horizontal axis = 0). Overall, prioritizing easy-to-geolocate images accelerates the geolocation time and enhances precision. The detail of the results is explained below.

We then asked experts to geolocate 39 images. Those images were classified using previously trained model. From that automated classification,about 22 images were predicted as Easy. Figure 13 shows that most of the images classified as Easy are geolocated with high precision (within a few meters). Most of the images geolocated with high precision were geolocated in less than 4 minutes. If we assume an average of 3 minutes per photo, a person could potentially geolocate 20 images per hour. Consequently, in a group of 10 volunteers, we could geolocate 200 images per hour with high precision. On the other hand, 17 images were predicted as Difficult. Figure 14 shows the precision and time spent on geolocating images classified as Difficult to geolocate. We could see how most of them are geolocated with low precision within the city level.

Although this is a preliminary result based on a limited number of images, it clearly shows that prioritizing images easy-to-classify during the crowdsourcing process could potentially augment the overall geolocation process, by increasing the precision and reducing the time needed to geolocate the images.

[12]https://en.wikipedia.org/wiki/Fleiss_kappa

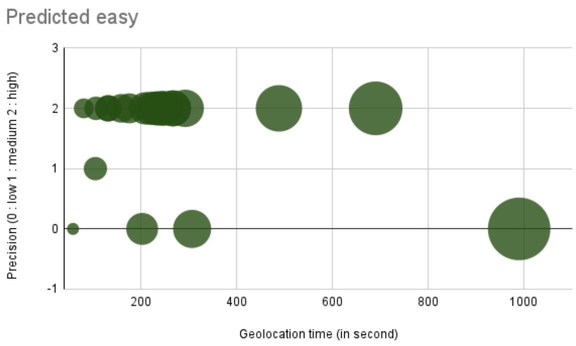

Figure 13. Geolocation level of images predicted as Easy to geolocate. The horizontal axis and the size of the radii represent the average time spent by users to geolocate the photo

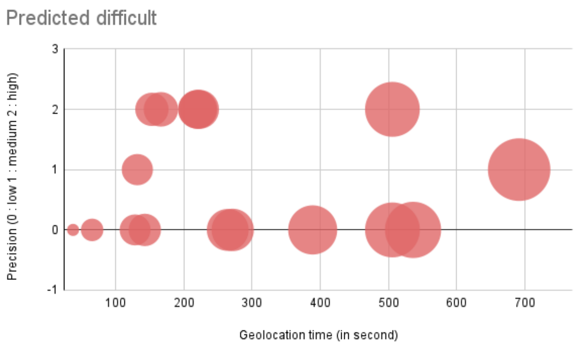

Figure 14. Geolocation level of images predicted as Difficult to geolocalize. The horizontal axis and the size of the radii represent the average time spent by users to geolocate the photo

CONCLUSIONS AND FUTURE WORK

Finding the location of social media images related to a disaster is a key process to make the data actionable, i.e., data that can be used to enable better-informed decisions. Geolocation of social media images is a task usually performed by humanitarian communities such as Stand By Task Force, GISCorp, or VOSTPT. It is a well-known fact that crowdsourcing the geolocation of social media images is always a difficult task which slows down the process of assessing the impact of a natural or man-made disaster in the immediate aftermath of the event. This preliminary research activity demonstrated that it is indeed possible to automatically predict the difficulty of geolocating an image, thus allowing to filter out impossible-to-geolocate images and giving priority to images that can be geolocated within a few minutes with high precision. Consequently, the proposed approach would enable a more efficient and timely analysis of social media data which is critical in a disaster.

Finally, regarding future work, first, the proposed approach needs to be evaluated in a real-time crisis response to demonstrate the potential of using social media data to assess the impact of disaster within 24 hours. Second, we will increase the number of geolocated images to answer the research questions in a robust manner.

ACKNOWLEDGEMENTS

This work was funded by the Indonesia Endowment Fund for Education (LPDP) and partially supported by the European Union's Horizon research and innovation programme Crowd4SDG and K-HealthInAIR projects (under grant agreements ID 872944 and 101057693, respectively) and by the grant PID2022-136787NB-I00 funded by MCIN/AEI/10.13039/501100011033.

REFERENCES

Alam, F., Ofli, F., & Imran, M. (2018). Crisismmd: Multimodal twitter datasets from natural disasters. *Proceedings of the 12th International AAAI Conference on Web and Social Media (ICWSM).*

Alam, F., Ofli, F., Imran, M., Alam, T., & Qazi, U. (2020). Deep learning benchmarks and datasets for social media image classification for disaster response. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 151–158. https://doi.org/10.1109/ASONAM49781.2020.9381294

Cerquides, J., & Mülâyim, M. O. (2022, January). Crowdnalysis: A software library to help analyze crowdsourcing results. https://doi.org/10.5281/zenodo.5898579

Cerquides, J., Mülâyim, M. O., Hernández-González, J., Ravi Shankar, A., & Fernandez-Marquez, J. L. (2021). A conceptual probabilistic framework for annotation aggregation of citizen science data. *Mathematics*, *9*(8). https://doi.org/10.3390/math9080875

Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 20–28. https://doi.org/10.2307/2346806

Firmansyah, H. B., Cerquides, J., & Fernandez-Marquez, J. L. (2022). Ensemble learning for the classification of social media data in disaster response. *ISCRAM 2022 Conference Proceedings – 19th International Conference on Information Systems for Crisis Response and Management*.

Firmansyah, H. B., Fernandez-Marquez, J. L., Cerquides, J., Lorini, V., Bono, C. A., & Pernici, B. (2023). Enhancing disaster response with automated text information extraction from social media images. *The 9th IEEE International Conference on Big Data Computing Service and Machine Learning Applications*.

González, D. L., alejandrodob, therealmarv, Keegan, M., Mendes, A., Pollock, R., Babu, N., Fiordalisi, F., Oliveira, N. A., Andersson, K., & et al. (2020). Scifabric/pybossa: V3.1.3. https://doi.org/10.5281/zenodo.3882334

Gopalan, R., Li, R., Patel, V. M., & Chellappa, R. (2015). https://doi.org/10.1561/0600000057

Han, B., Cook, P., & Baldwin, T. (2012). Geolocation prediction in social media data by finding location indicative words. *Proceedings of COLING 2012*, 1045–1062.

Huang, B., & Carley, K. M. (2020). A large-scale empirical study of geotagging behavior on twitter. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 365–373. https://doi.org/10.1145/3341161.3342870

Lorini, V., Panizio, E., & Castillo, C. (2022). Smdrm: A platform to analyze social media for disaster risk management in near real time. *In Workshop Proceedings of the 16th International AAAI Conference on Web and Social Media*.

Lorini, V., Rufolo, P., & Castillo, C. (2022). Venice was flooding ... one tweet at a time. *Proc. ACM Hum.-Comput. Interact.*, *6*(CSCW2). https://doi.org/10.1145/3555107

Middleton, S. E., Kordopatis-Zilos, G., Papadopoulos, S., & Kompatsiaris, Y. (2018). Location extraction from social media: Geoparsing, location disambiguation, and geotagging. *ACM Trans. Inf. Syst.*, *36*(4). https://doi.org/10.1145/3202662

Murgese, F., Costa-Alcaina, G., Mülâyim, M. O., Cerquides, J., & Fernandez-Marquez, J. L. (2022). Automatic outdoor image geolocation with focal modulation networks. *Artificial Intelligence Research and Development: Proceedings of the 24th International Conference of the Catalan Association for Artificial Intelligence*.

Ofli, F., Alam, F., & Imran, M. (2020). Analysis of social media data using multimodal deep learning for disaster response. *arXiv preprint arXiv:2004.11838*.

Ravi Shankar, A., Fernandez-Marquez, J. L., Pernici, B., Scalia, G., Mondardini, M. R., & Serugendo, G. (2019). Crowd4EMS: A crowdsourcing platform for gathering and geolocating social media content in disaster response. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-3/W8*, 331–340. https://doi.org/10.5194/isprs-archives-xlii-3-w8-331-2019

Rufolo, P., Muraro, D., & Lorini, V. (2021). *Social media image analysis in the immediate aftermath of the 2020 beirut blast*.

Scalia, G., Francalanci, C., & Pernici, B. (2022). CIME: Context-aware geolocation of emergency-related posts. *GeoInformatica*, *26*(1), 125–157. https://doi.org/10.1007/s10707-021-00446-x

Tahmasebzadeh, G., Hakimov, S., Ewerth, R., & Müller-Budack, E. (2023). Multimodal geolocation estimation of news photos. In J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, & A. Caputo (Eds.), *Advances in information retrieval* (pp. 204–220). Springer Nature Switzerland.

Utomo, M. N., Adji, T. B., & Ardiyanto, I. (2018). Geolocation prediction in social media data using text analysis: A review. *2018 International Conference on Information and Communications Technology (ICOIACT)*. https://doi.org/10.1109/icoiact.2018.8350674

Venkatagiri, S., Thebault-Spieker, J., Kohler, R., Purviance, J., Mansur, R. S., & Luther, K. (2019). Groundtruth: Augmenting expert image geolocation with crowdsourcing and shared representations. *Proc. ACM Hum.-Comput. Interact.*, *3*(CSCW). https://doi.org/10.1145/3359209

Xie, Q., Luong, M.-T., Hovy, E., & Le, Q. V. (2020). Self-training with noisy student improves imagenet classification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–10695. https://doi.org/10.1109/CVPR42600.2020.01070