

Supplementary Materials - LADI v2

April 2025

Author Statement

The authors of this scientific paper bear full responsibility for any violation of rights that may arise from the collection of the data included in this research. LADI v2 is licensed under the Creative Commons-Attribution 4.0 for data and MIT license for code.

Contents

1 Datasheet for Dataset LADI v2	1
1.1 Motivation	1
1.2 Composition	2
1.3 Collection process	4
1.4 Preprocessing/cleaning/labeling	5
1.5 Uses	6
1.6 Distribution	7
1.7 Maintenance	7
2 Data Collection Details	8
2.1 Labeling Campaign	8

1 Datasheet for Dataset LADI v2

Based on Datasheets for Datasets [1]

1.1 Motivation

The questions in this section are primarily intended to encourage dataset creators to clearly articulate their reasons for creating the dataset and to promote transparency about funding interests.

For what purpose was the dataset created?

This dataset was created to fill a gap in the availability of high-quality annotated low-altitude aerial imagery of areas affected by natural disasters. We hope that it will be used to train algorithms to automate the process of identifying useful images of interest captured in response to natural disasters and facilitate the work of emergency managers.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

Who funded the creation of the dataset?

The images in the dataset were captured by Civil Air Patrol (CAP) volunteers between 2015 and 2023. They were labeled by CAP volunteers in December 2023. The technical work of curating the dataset, conducting the annotation campaign, quality control, and training reference baseline classifiers was performed and funded by ANONYMIZED FOR REVIEW.

1.2 Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?

Each example in the dataset is an RGB image of a site affected by a disaster. The images were captured from low-flying planes by Civil Air Patrol volunteers, using either handheld cameras or aircraft-mounted systems such as the WaldoAir system. The images have timestamps and geotags in their EXIF metadata. Each image is labeled for the presence of 15 classes described below.

How many instances are there in total (of each type, if appropriate)?

The dataset contains 9,963 examples, divided into a train split of 8030 examples, a validation split of 893 examples, and a test split of 1041 examples.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

The dataset is a subsample of all images collected by CAP between 2015-2023. The validation and train sets consist of images gathered between 2015-2022, and the test set consists of images gathered during 2023. The images were sampled randomly without replacement and are likely representative of the disasters which took place during this time.

What data does each instance consist of?

Each example in the dataset is an RGB image of a site affected by a disaster. The images were captured from low-flying planes by Civil Air Patrol volunteers, using either handheld cameras or systems such as the WaldoAir system.

Is there a label or target associated with each instance?

Each image is labeled for the presence of the 15 classes described below:

- bridges_any
- bridges_damage
- buildings_affected
- buildings_any
- buildings_destroyed

- buildings_major
- buildings_minor
- debris_any
- flooding_any
- flooding_structures
- roads_any
- roads_damage
- trees_any
- trees_damage
- water_any

In v2a of the dataset, we removed the “bridges_any” label because there were not enough instances of damaged bridges, and combined the “buildings_destroyed”, “buildings_major”, “buildings_minor”, and “buildings_affected” labels, which are based on FEMA Preliminary Damage Assessment (PDA) guidelines, into “buildings_minor_or_greater” and “buildings_affected_or_greater”, giving the following label set:

- bridges_any
- buildings_any
- buildings_affected_or_greater
- buildings_minor_or_greater
- debris_any
- flooding_any
- flooding_structures
- roads_any
- roads_damage
- trees_any
- trees_damage
- water_any

As several of these elements are expected to be present in many of the images, this dataset is labeled with multi-label classification in mind.

Is any information missing from individual instances?

Images occasionally contain watermarks or timestamps which can obscure image contents.

Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)?

No

Are there recommended data splits (e.g., training, development/validation, testing)?

Yes; the splits are described above. The test split is intentionally drawn from a different timeframe from the training and validation sets, to simulate the distribution shift expected when training on historical data and then deploying the model for future data.

Are there any errors, sources of noise, or redundancies in the dataset?

The images are taken from operational data and are meant to capture specific challenges of working with data of this sort. As a result some images are uninformative. Some notable examples include images of the planes in the airfields before takeoff, or the interior of the plane. Some images are redundant in that they capture the same area of interest from similar perspectives. Some of the images are taken in poor lighting conditions, or from a perspective or angle that makes it difficult to resolve features of interests. Some of the images contain parts of the plane (struts or wings) which obscure useful content. We chose to include all of these types of images in the dataset because it is representative of the kind of noise that occurs with this kind of data in practice.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?

The dataset is self-contained. It is a partial copy of the `fema-cap-imagery` Amazon S3 bucket and is itself hosted on S3 under the Amazon Open Data Sponsorship Program. A copy of the smaller (resized) version of the dataset is hosted on Hugging Face.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)?

No. The dataset contains images of people's homes, but the images are captured from public airspace.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?

The dataset contains images of destroyed homes and neighborhoods.

Does the dataset relate to people?

No.

1.3 Collection process

How was the data associated with each instance acquired?

The images in the dataset were captured during the course of CAP missions and labeled in triplicate by trained CAP volunteers. The labeling in triplicate enabled volunteers to verify each other's work.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?

The data was copied from the public `fema-cap-imagery` s3 bucket, and filtered to include only images that were taken for federally declared disasters between 2015-2023. This was done by filtering for images whose location and timestamp metadata were taken within a county with a federal disaster declaration whose start date was no more than 14 days earlier.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

The sampling was uniformly random without replacement from the set of images associated with federal disaster declarations. The validation and train sets consist of images gathered between 2015-2022, and the test set consists of images gathered during 2023.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

The CAP volunteers were not compensated for their time.

Over what timeframe was the data collected?

The images were captured between 2015 and 2023. They were labeled in December 2023.

Were any ethical review processes conducted (e.g., by an institutional review board)?

No

Does the dataset relate to people?

No

1.4 Preprocessing/cleaning/labeling

The questions in this section are intended to provide dataset consumers with the information they need to determine whether the “raw” data has been processed in ways that are compatible with their chosen tasks. For example, text that has been converted into a “bag-of-words” is not suitable for tasks involving word order.

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?

Images were de-duplicated based on their md5 checksum. Images were annotated by three distinct teams, and final labels were assigned by majority vote.

Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?

The individual responses from the annotators are saved privately.

Is the software used to preprocess/clean/label the instances available?

The specific software for de-duplication used is not available; semantically identical software can be easily reproduced as a single line of Linux shell script if desired.

The interface for labeling is not made public since it requires authentication using FEMA credentials.

Any other comments?

1.5 Uses

These questions are intended to encourage dataset creators to reflect on the tasks for which the dataset should and should not be used. By explicitly highlighting these tasks, dataset creators can help dataset consumers to make informed decisions, thereby avoiding potential risks or harms.

Has the dataset been used for any tasks already?

We have trained baseline multi-label classifier models for this dataset.

Is there a repository that links to any or all papers or systems that use the dataset?

The code to train our models is available at: ANONYMIZED FOR REVIEW

What (other) tasks could the dataset be used for?

The data could be used as a pretraining task for another project using aerial imagery. The labels may be used to generate captions for training vision-language models. The data may also be used to support autonomy tasks for UAVs.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

The data is taken from federally declared disasters in the United States in 2015-2023. As such, the data is only representative of the disaster types and locations present in the data. As climate change and human development affect the distribution and severity of disasters and its impacts, the representativeness of the data may change. For example, regions that have not historically experienced frequent hurricanes may experience them more frequently in the future. Performance may degrade in such situations.

Are there tasks for which the dataset should not be used?

We believe a great deal of caution should be exercised in attempting to identify the subjects of this data. Though the imagery and metadata in this dataset was publicly available before we released it with labels, we encourage anyone who uses this dataset to respect the privacy of disaster victims.

Any other comments?

1.6 Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?

The dataset is publicly available.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?

The dataset is available for direct download on s3: ANONYMIZED FOR REVIEW or can be conveniently loaded as a Hugging Face dataset: ANONYMIZED FOR REVIEW

When will the dataset be distributed?

The dataset is currently available.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

The dataset is available under Creative Commons Attribution 4.0 (CC-BY-4.0), and the code is available under the MIT license.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

No

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

No

Any other comments?

1.7 Maintenance

These questions are intended to encourage dataset creators to plan for dataset maintenance and communicate this plan with dataset consumers.

Who is supporting/hosting/maintaining the dataset? How can the manager of the dataset be contacted (e.g., email address)?

The dataset is hosted on Amazon S3 and Hugging Face, and maintained by ANONYMIZED FOR REVIEW

Is there an erratum?

No

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?

There are no current plans to update the original dataset. However, it is permissively licensed and we encourage users to create forks as they desire.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)?

N/A

Will older versions of the dataset continue to be supported/hosted/maintained?

v1 of the dataset continues to be hosted on S3, and the code for this version maintains compatibility for a subset of tasks on v1.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

Updates to the dataset can be proposed on the project GitHub page: ANONYMIZED FOR REVIEW Updates to the dataset will appear in the README on the main page. Anyone is free to fork the dataset for any reason.

2 Data Collection Details

2.1 Labeling Campaign

In collaboration with the CAP Geospatial Program, volunteers were recruited from a group of CAP volunteers previously trained to assist FEMA in conducting image-based residential damage assessments. 46 participated in the campaign.

Volunteers were divided into three groups of nineteen, each led by a member of the CAP Geospatial Program Leadership Team and designated by color (Yellow Team, Green Team, Blue Team). Each group was assigned a web application, designed in ArcGIS Online, for annotating images. In these applications, volunteers could "claim" images to annotate as a point on a map and view the available, work in progress, and completed annotations. The same images were shown to each group, ensuring each image was reviewed by three different volunteers. An example of the interface that the annotators used is given in Figure 1.

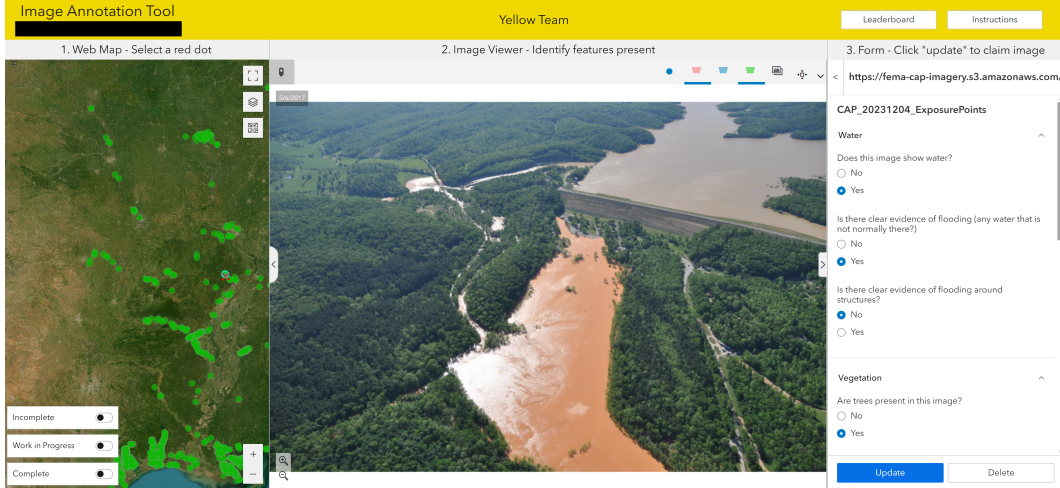


Figure 1: Example of the annotation interface for the Yellow Team

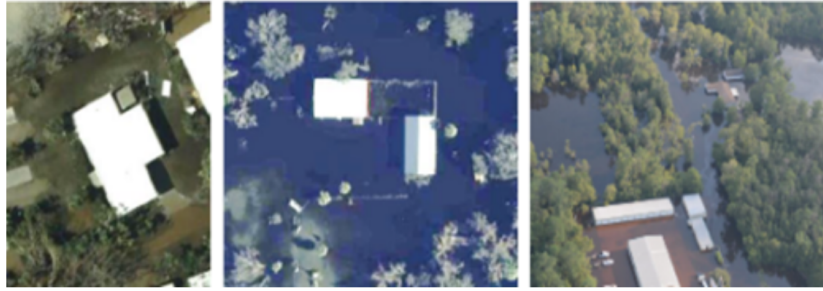
For each image, volunteers were asked a series of cascading yes/no questions indicating the presence or absence of features in the images. An open text field allowed volunteers to describe additional features of interest or comment on the quality of the image. The details of the instructions are provided in Table 1. For privacy, the “editor” field is not included in the released dataset. The “notes” field is also not released in this version of the dataset. The released dataset aggregates the responses from the three annotators for each image and provides the majority-vote label for each of the categories.

We instructed that the annotators mark that an image contained a feature only if it was *clearly* visible in the image; if the annotator was unsure whether it was clearly visible, we recommended that they mark it as “no”. We provided image references for the various building damage levels, as shown in Figure 2.

The campaign was conducted in two rounds between December 19th and 30th, 2023. Volunteers labeled a minimum of 8 and maximum of 2544 images each, totalling an estimated 250 person-hours. Volunteers used Microsoft Teams to ask questions, identify technical glitches, and provide motivation. Volunteers were mailed letters and tokens of appreciation after the campaign.

References

- [1] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. I. au2, and K. Crawford, “Datasheets for datasets,” 2021.



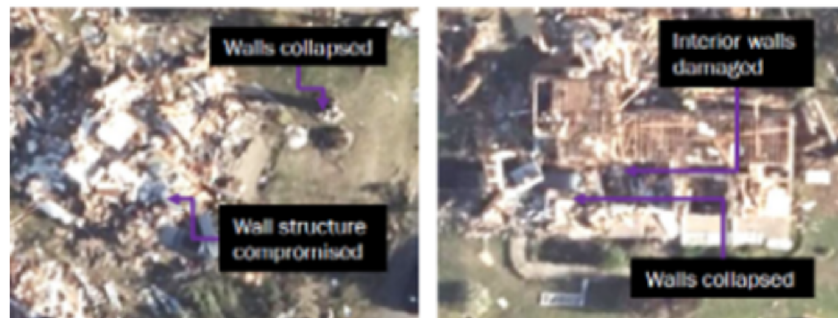
(a) Examples of affected buildings



(b) Examples of minorly damaged buildings



(c) Examples of majorly damaged buildings



(d) Examples of destroyed buildings

Figure 2: Image examples for damage levels

Table 1: Annotation fields and instructions

Field Name	Instructions/Details
URL	Image URL
editor	Claim this image
water_any	Does this image show water of any sort?
flooding_any	Is there clear evidence of flooding (any water that is not normally there?)
flooding_structures	Is there clear evidence of flooding around structures?
trees_any	Are trees present in this image?
trees_damage	Are any trees damaged? (uprooted, downed, flooded, burned)
debris_any	Is there debris in this image?
buildings_any	Are buildings clearly visible?
buildings_affected	Does this image clearly show any affected buildings?
buildings_minor	Does this image clearly show any buildings with minor damage?
buildings_major	Does this image clearly show any buildings with major damage?
buildings_destroyed	Does this image clearly show any destroyed buildings?
roads_any	Are any roads clearly visible?
roads_damage	Does this image clearly show damaged roads? (flood, pavement damage/washout, debris)
bridges_any	Are any bridges clearly visible?
bridges_damage	Does this image clearly show damaged bridges? (flood, pavement damage/washout, debris)
notes	Please describe any other useful information this image provides.