Download the data

Overview

The xView2 xBD dataset is split into four sets: train, test, holdout, and tier3. For the xView2 Challenge, the training set provided pairs of images (pre- and post-disaster) and ground truth information about buildings and damage levels for the pixelwise segmentation task; the test set provided images only and ground truth targets were used to provide a public leaderbooard, and the holdout set was kept private for the purpose of testing generalization performance on verified Challenge submissions. The Tier3 data became available mid-way through the Challenge, and were provided as an additional/ supplementary training set, covering additional span of disasters and geographical area.

Now that the xView2 Challenge is complete, we are making each set available for download. There was great community engagement throughout xView2, but the Challenge itself is just the first chapter of an ongoing story where the AI/ML community uses data like these to positively impact HADR missions. High quality datasets are an important part of applied research and innovation, and we look forward to seeing what the community will continue to do with the xView2 data!

During the Challenge, ground truth was provided in the same format we specified for predictions, for the Challenge task. With this public release, in addition to those ground truth prediction targets, we are also releasing additional label information, including polygons and some additional metadata (e.g., view parameters). We believe this information will be helpful to the community, especially on projects where a different task or specification can benefit from the new metadata.

Training sets

The "train" dataset contains 2799 pairs of high-resolution RGB satellite imagery in PNG format, provided as a 7.8 GB GZIP archive. The datasets are organized into pairs of images. Each pair is a training instance: one image from before the natural disaster (e.g., `socal-fire_00001390_pre_disaster.png`), and another image from after a disaster occurred (e.g., `socal-fire_00001390_post_disaster.png`). Filenames in the training set indicate the location and type of disaster, a numerical ID for the pair of images, and whether the image is "pre" or "post" disaster event. After you unpack the archive, you should see three directories: images, labels, and

targets. Images are RGB PNG images. Labels are JSON files with building annotations (polygons in WKT format) and metadata for the given view. Targets are pixelwise ground truth in the format specified for the xView2 Challenge, which were derived from the underlying polygons.

Test and Holdout sets

The test and holdout datasets each include 933 instances. Label information (polygon annotations and metadata) and ground truth targets (pixelwise values in PNG format) are also now available for Test and Holdout sets. During the Challenge, filenames for the test set were obfuscated to prevent leakage of disaster type and location; now the filenames follow the same pattern as the training set.

License

Before Downloading...

Please make sure you have the required storage space available, review the Data Use Terms, and if you agree then click the button below to initiate a download. You will be redirected to download links that are only valid for a limited time; if your download is interrupted, return to this page and click the button below to obtain fresh links. Please be respectful of download bandwidth, and re-download data only as necessary. Do not distribute download links or embed download functionality into online services or cloud instances. If you wish to share the dataset, direct users to xview2.org to sign up, review terms, and download.

By downloading the data, you agree to the Terms of Data Use