## Plan Overview

*A Data Management Plan created using DMPTool*

**DMP ID:** https://doi.org/10.48321/D10S4D

**Title:** Two-stage Ontology-Guided Attribute Learning for Zero-Shot or Low-Shot Classification Using Coincidentally Collected Data

**Creator:** Hui Wang - **ORCID:** 0000-0002-9412-3516

**Affiliation:** Florida State University (fsu.edu)

**Principal Investigator:** Hui Wang

**Data Manager:** Hui Wang, Hongmei Chi

**Contributor:** Hongmei Chi

**Funder:** United States Air Force Research Laboratory (wpafb.af.mil)

**Funding opportunity number:** POA-23-RI-002

**Template:** Digital Curation Centre

**Project abstract:**

The goal of this project is to establish a new methodology, ontology-guided attribute learning, to fuse the information in multiple data modalities to assist low-shot or zero-shot classification for object recognition. The methodology is built under a self-supervised learning (SSL) framework, by which stage-1 learning pre-trains ontology about class attributes from multi-modal data, followed by fine-tuning ontology and class embedding in stage-2 learning. The class embedding and ontology learned from data provide attributes of target classes of interest. In downstream learning tasks, the attributes of classes allow for object recognition based on limited or zero observations in training data by mimicking human recognition of new objects via identifying a unique combination of attributes about the targets. Training and education opportunities will be created for 2 graduate and 2 undergraduate students. The project aligns with AFRL's interest in reconstruction and recognition based on new data modalities, such as Synthetic Aperture Radar (SAR) and electro-optical (EO) images, along with optical images, texts, and/or computer-aided (CAD) models coincidentally collected for the same objects.

**Start date:** 08-21-2023

**End date:** 08-20-2025

**Last modified:** 05-10-2023

**Copyright information:**



---

## Two-stage Ontology-Guided Attribute Learning for Zero-Shot or Low-Shot Classification Using Coincidentally Collected Data

## Data Collection

---

### What data will you collect or create?

The data will be collected from public datasets published online, including (Please refer to the details in Task 1 in the proposal.)

1. UNified COincident Optical and Radar for recognitioN (UNICORN) 2008 dataset, including SAR/EO NITF image data, Ground Truth using GPS, and WAMMIT Truth database for EO data.

*https://github.com/AFRL-RY/data-unicorn-2008*

2. Stationary Target Acquisition and Recognition (MSTAR) Dataset, including SAR images and facet files (CAD models) of different types of vehicles.

*https://www.sdms.afrl.af.mil/index.php?collection=mstar*

3. Supplementary dataset. Supplementary information on the objects in the two datasets above will be collected from other sources, including ImageNet containing optical images of vehicles, manufacturers' specifications, and descriptions of vehicles.

Data will be created during the development of algorithms and use cases, including

1. Pre-training, training, and testing datasets for object recognition integrating SAR/EO/optical images with text information from labels/manuals/literature, and/or CAD models (facet files).

2. Intermediate results generated during the development of algorithms and use case development. These data are organized by .csv, .mat, or .txt files for matrices, Matlab (.m)/Python files including Jupyter Notebook (.py, .ipynb) files, plots or synthetically generated images saved as .png/jpg/tif or .fig files

These data formats can enable sharing and long-term access.

### How will the data be collected or created?

Data collection will be performed by directly visiting the open-access webpage and GitHub for downloading, sending written permission requests to the dataset organizers, and paying access fees for cloud datasets published on AWS.

Data will be created by storing intermediate results during algorithm development, reporting visualized results and algorithm performance, comparison of results among different methods, documentation by MS Word for paper publication, and Matlab/Python codes.

Codes, results, and documents will be organized under different older on Dropbox. All historical versions of

codes and generated results will be organized under different subfolders with the creation date and modifier's name included in the folder name on Dropbox. The document will be co-edited by PI/co-PI and students.

The data quality is ensured by data cleaning and pre-processing, including the elimination of outliers, filtering of noisy data, inspection for erroneous records, and organization of coincidental data by overlapping or matching information based on specific criteria, such as location and time.

## Documentation and Metadata

### What documentation and metadata will accompany the data?

The documentation includes tabular and visualized results/plots, verbal explanations of the results in .txt files or MS word documents/Excel sheets, comments in the code file (.m, .py, .ipynb), and documents of paper writing in different versions. The documentation will also include debugging issues, including types of errors encountered and how the problems are solved.

The metadata includes the description of parameter settings/data used, personnel in charge, and date/time for each algorithm implementation and debugging (instance). The documentation and metadata of each data experiment will be created right after the analysis by the team members responsible for the task (PI/Co-PI and students).

## Ethics and Legal Compliance

### How will you manage any ethical issues?

The data in this project are all from publicly available data online or requests with permission by contacting the data manager. There are no privacy issues with the data. No sensitive or classified data are included in this project.

### How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The raw data downloaded from web pages are publicly available. The data generated in this project will be jointly owned by all the PI/co-PI and AFRL. There is no restriction on the reuse of third-party data. With the AFRL's approval, the project team will publish selected data, including codes, on the PI's Github or Zenodo account. The approval process for data publication can take several months.

## Storage and Backup

### How will the data be stored and backed up during the research?

Excessively large data downloaded will be saved on external drives at the PI/co-PI's labs. Other data will be saved on the PI's cloud drive using his enterprise-version Dropbox subscribed by Florida A&M University. He will create shared folders with the co-PI and students. The project team will utilize the historical version

restoration function of Dropbox to deal with accidental data losses. In addition, extra data storage devices will be purchased to save the data as backup, including portable solid-state drives and/or network-attached storage.

### How will you manage access and security?

All key members in the project (PI, co-PI, graduate students, and undergraduate students receiving support from this project, along with AFRL collaborators) have equal editing access to the data on Dropbox and raw data downloaded to the storage in the PI/co-PI's lab. A potential risk to data security is unauthorized data sharing with external members violating ITAR. To manage the data security, edit access will only be granted to key members as secured by Dropbox sharing functions. Data storage on SSD and NAS will be strictly limited to project team members with a passcode and will not be shared externally.

## Selection and Preservation

---

### Which data are of long-term value and should be retained, shared, and/or preserved?

The data that are considered to have long-term values include: (1) All the results from successful implementations of the algorithms in use cases, (2) Finalized error-free codes with annotations added, and (3) Organized and cleaned input datasets (pre-training, training, testing sets). Data selection is decided by how they contribute to developing case studies and debugging.

These data will help future scholars to repeat the results and identify opportunities for improvement. Meanwhile, they provide guidelines and lessons learned from this process for future developers so that algorithm development efficiency can be improved.

There is no time limitation on long-term data preservation after AFRL's approval.

### What is the long-term preservation plan for the dataset?

All the long-term data will be saved on the Dropbox cloud drive with an enterprise-user subscription by Florida A&M University. The data preservation should be subject to the International Traffic in Arms Regulations ("ITAR," 22 CFR 120-130).

With the AFRL's permission, some selected long-term data for open access will be uploaded to (1) the PI's Github/Zenodo account, (2) Research journal websites that support electronic supplemental materials providing permanent data access and preservation, and (3) a virtual repository service offered by FAMU-FSU College of Engineering called *DigiNole Commons* that allows materials deposit, access, and preservation. For university users, these preservation methods are all free of charge.

## Data Sharing

---

### How will you share the data?

Subject to export control, permission for data access from domestic external users will be obtained by both PI

and AFRL. Then, dedicated subfolders will be created, and selected data will be deposited in these folders to be shared with external users via Dropbox. External domestic users need to abide by ITAR regulations and only have "Read" access to these data and cannot modify them on the cloud drive (Dropbox).

The data can be available after the completion of this project. With the AFRL's approval, selected data may be shared at the time of paper publication. In addition, through the journal websites that support electronic supplemental materials, the project team will publish the selected dataset and/or links with the PI's Github or Zenodo account.

### Are any restrictions on data sharing required?

The data will be subject to export control to foreign entities as specified by ITAR. This research is subject to publication and foreign national restrictions. No part of this research can be shared with foreign nationals or published without the written approval of AFRL. The project team will share the data without violating the restriction.

## Responsibilities and Resources

---

### Who will be responsible for data management?

The PI and co-PI will be equally responsible for all the data organization, access/management policies, and sharing under the restriction of ITAR. There will be no change to the scope of this data management plan should any of the PIs leave their institutions.

### What resources will you require to deliver your plan?

All the data will be uploaded to the enterprise Dropbox account with the university subscription, which is offered for faculty for free. Without PI/co-PI and ARFL's permission, the IT at Florida A&M University and Dropbox will not disclose the information to anyone outside the project team.

All new students who join the project team will receive training by the PI or co-PI on data access and preservation, as outlined above.

---

## Planned Research Outputs

### Dataset - "Multi-modal dataset for low-shot/zero-shot classification of objects"

The dataset includes pre-training, training, and testing datasets for object recognition based on the integration of SAR/EO/optical images and text coincidentally collected on the same objects.

### Software - "Automated ontology generation tool for characterizing target objects "

A software package consisting of a set of algorithms to automatically generate ontology characterizing the object of interest, given the input of SAR/EO/optical images and text from literature/manual.

### Software - "Zero-shot or low-shot classification tool "

A software package performing zero-shot and low-shot classification to recognize objects from SAR images while integrating with EO/optical images and text information.

### Use cases for demos - "Use cases for zero-shot or low-shot object recognition "

Use cases include codes, pre-training/training/testing datasets, and metadata files that explain how the algorithms are implemented. Examples will be provided in a document file to illustrate the procedure.

### Data paper - "Journal and/or conference paper(s) on low-shot/zeros-shot object recognition by integrating multi-modal data"

Papers are to be submitted to conferences and academic journals in the IEEE and INFORMS communities with the AFRL's approval.

---

## Planned research output details

| Title | Type | Anticipated release date | Initial access level | Intended repository(ies) | Anticipated file size | License | Metadata standard(s) | May contain sensitive data? | May contain PII? |
|---|---|---|---|---|---|---|---|---|---|
| Multi-modal dataset for low-shot/zero-shot classif ... | Dataset | Unspecified | Restricted | None specified | 1 TB | None specified | None specified | No | No |
| Automated ontology generation tool for characteriz ... | Software | Unspecified | Restricted | None specified | | None specified | None specified | No | No |
| Zero-shot or low-shot classification tool | Software | Unspecified | Restricted | None specified | | None specified | None specified | No | No |
| Use cases for zero-shot or low-shot object recogni ... | Use cases for demos | Unspecified | Restricted | None specified | | None specified | None specified | No | No |
| Journal and/or conference paper(s) on low-shot/zer ... | Data paper | 2025-12-30 | Open | None specified | | None specified | None specified | No | No |