**Phase II Sample Dataset Documentation**

*Dataset Description*

The SeabedObjects open source dataset contains categorized side-scan sonar images that describe submerged ships, submerged airplanes, and empty seafloor. Selected examples of each category of image are provided in /Seabed\_Objects\_Demo/Overview\_of\_Seabed\_Images.ipynb.

The images are then funneled into three datasets, each labeled with the classes “plane”, “seafloor”, and “ship”.

1. /Seabed\_Objects\_Demo/unbalanced\_training\_validation\_set

Contains all of the whole, most-detailed planes (38), the clearest, most-detailed large ships (239), and a randomly-selected subset of empty seafloor (378).

1. /Seabed\_Objects\_Demo/balanced\_training\_validation\_set

A more-balanced subset of the first dataset (described in the Available Models below).

1. /Seabed\_Objects\_Demo/test\_set

A “new environment” containing the remaining seafloor images (199), the small and limited-detail ships (138), and the fractured and limited-detail planes (24).

*Available Models*

The notebook establishes two models: a model trained on all the available images in the training dataset, and a model trained on a much smaller subset of the dataset that balances the three classes more evenly to reduce bias. We have provided the following saved versions of the two models:

1. unbalanced\_model.pth
2. balanced\_model.pth

The current notebook setup for each model can be found below:

**Unbalanced Model Balanced Model**

***Trained & Validated on:*** *Planes* 38 38

*Seafloor* 378 50

*Ships* 239 50

***Hyperparameters:*** *Image resize*  150 x 150 150 x 150

*Batch size* 100 20

*Epochs* 20 20

*Learning rate* 0.001 0.0005

*Functionality to Demonstrate (Suggested)*

Below, we provide additional detail on how you may wish to use the demo dataset and models to showcase your platform according to the desired capabilities listed in the original MLOPs solicitation. Note that while we have provided two baseline dataset partitions (balanced and unbalanced) and models trained on those dataset partitions, we look forward to seeing Phase II presentations that expand on our very basic work. Note that what we have provided incorporates no hyperparameter tuning, dataset cleaning / versioning, advanced augmentation, large-scale experiments, deployment steps, T&E, observability, or OOD detection. We leave it to your team to expand on what we have provided to best demonstrate the capabilities of your platform.

Management of data provenance and quality

To do this using the demo dataset, we suggest using the fact that we have already provided two different training dataset partitions (balanced and unbalanced) to demonstrate how data provenance and data quality are tracked and assessed. We also suggest potentially leveraging the deeper analysis of data samples provided in the “Overview of Seabed Images.ipynb” notebook to demonstrate data quality assessment capabilities.

Provide automation of dataset preparation activities (i.e. data cleaning, transformation, low latency data retrieval and dataset versioning and documentation) and/or ability to integrate with industry-standard data management tooling

Use the two different data partitions we have provided to demonstrate this. We would also be interested in any additional cleaning and transformation that your team wishes to show / demonstrate.

Experiment tracking, hyperparameter search metadata/results

We have done very little in terms of hyperparameter searches / extensive experiments to optimize performance (or architecture). If this is part of your solution, demonstrate how experiments can be started, tracked, and evaluated.

Model management, version tracking, storage, and CI/CD

At the least, we suggest ingesting the two models we have provided to demonstrate how model versions are managed, tracked, stored, and run through CI/CD pipelines for deployment.

Provide a model catalog, including the ability to search and access modeling assets developed within the solution by attribute (date developed, name, type, training dataset used etc.) and/or ability to easily integrate with industry-standard model repositories

We suggest at least using the two models provided for this. Ideally, your team could run a larger set of experiments using the dataset provided and demonstrate a cataloging ability with a larger number of models. One way to do this might be to run basic neural architecture search and use the results to populate a catalog . Another might be to have multiple developers build models and use each of their contributions.

Model observability

As described in the dataset “Overview of Seabed Images.ipynb” notebook referenced above, there exist a number of subtleties to this side-scan sonar dataset, which represents the exact modality to be used by the end customer for this CSO. We suggest using our preliminary analysis to get some ideas, and to use your platform’s observability to techniques to e.g. identify potential errors and characterize model performance in a more fine-grained way than simple performance metrics.

OOD and data drift detection.

There is known drift between the training and validation / test sets we have provided. Demonstrating how your platform might identify and surface this to the user – particularly at test time - would be valuable.

Allow DoD to Extract in-platform Data, Models, and Management Information in Open Formats

Demonstrate the process of ingesting the data and models we have provided into the platform. Similarly, demonstrate how the outputs of your analysis above - any experiments, hyperparameters, models, datasets / versions, CI/CD test – can be exported to open formats.

CI/CD for models deployed to IOT / edge devices

The simplest way to do this would be to use the two models we have provided and show (a) how one would switch from deploying one to deploying the other and (b) what the CI/CD process for that looks like. How is it confirmed that tests pass? How are tests written? Does the platform provide automated checks before deployment? Can the above be done from a deployed (e.g. on-prem) system versus a cloud deployment?

Containerized Development & Hardware / Platform Agnostic

Discuss how your platform supports (a) containerized installation / development and (b) how it may be deployed on both cloud and on-prem environments.

Deployment in IL5 Environment

Not needed for demonstration; environment will be unclassified. Can describe any existing deployments.