

Overview of Demo

Given:

Clear and easily labeled training/validation sets to tune a model for a CV classification task (classes=ship, plane, empty seabed)

Training/Validation Data (80/20 random split)

- 239 Large Ships (military, logistic, etc)
- 38 Planes (single-engine, multi-engine, military, etc)
- 378 Empty Seabed/Seafloor images (no manmade objects in FOV)

Implement MLOps concepts and capabilities:

Continue the ML lifecycle and update

Answers the following, e.g.:

- How is the model retrained
- What level of autonomy is present in the system
- How is performance presented to the user/how is degradation flagged
- How easy is end-user use and interface with the system

Train and Validate a model:

Achieve a nominal, acceptable level of performance (>95% e.g.)

Problems to address

- Imbalanced dataset, model bias
- Tuning
- Automated/User-friendly data pipeline
- Clear performance metrics

Assess model under test conditions

Possible issues that arise:

- Unacceptable performance degradation
- Incongruous data
- The need to retrain the model
- Contested/degraded operations

Enter a "new environment":

Ingest test data and assess model

Test Data

- 138 Small or Unclassifiable Ships
- 24 Blurred or Fractured Planes
- 199 New Empty Seabed/Seafloor images

What DIU provides to Vendors in Box: Option 1



Given:

Clear and easily labeled training/validation sets to tune a model for a CV classification task (classes=ship, plane, empty seabed)

Training/Validation Data (80/20 random split)

- 239 Large Ships (military, logistic, etc)
- 38 Planes (single-engine, multi-engine, military, etc)
- 378 Empty Seabed/Seafloor images (no manmade objects in FOV)

A folder with 3 labeled subfolders each with images as stated (optional: balanced dataset ~50/38/50 split)

Implement MLOps concepts and capabilities: Continue the ML lifecycle and update

Answers the following, e.g.:

- How is the model retrained
- What level of autonomy is present in the system
- How is performance presented to the user/how is degradation flagged
- How easy is end-user use and interface with the system

Train and Validate a model:

Achieve a nominal, acceptable level of performance (>95% e.g.)

Problems to address

- Imbalanced dataset, model bias
- Tuning
- Automated/User-friendly data pipeline
- Clear performance metrics

A trained and validated model in a .ipynb notebook

Assess model under test conditions

Possible issues that arise:

- Unacceptable performance degradation
- Incongruous data
- The need to retrain the model
- Contested/degraded operations

Test set results included in .ipynb notebook with saliency maps and T-SNE plots

Enter a "new environment":

Ingest test data and assess model

Test Data

- 138 Small or Unclassifiable Ships
- 24 Blurred or Fractured Planes
- 199 New Empty Seabed/Seafloor images

A folder with 3 labeled subfolders each with images as stated

Rojo and Wreck stop here

What DIU provides to Vendors in Box: Option 2



Given:

Clear and easily labeled training/validation sets to tune a model for a CV classification task (classes=ship, plane, empty seabed)

Training/Validation Data (80/20 random split)

- 239 Large Ships (military, logistic, etc)
- 38 Planes (single-engine, multi-engine, military, etc)
- 378 Empty Seabed/Seafloor images (no manmade objects in FOV)

A folder with 3 labeled subfolders each with images as stated (optional: balanced dataset ~50/38/50 split)

Implement MLOps concepts and capabilities:

Continue the ML lifecycle and update

Answers the following, e.g.:

- How is the model retrained
- What level of autonomy is present in the system
- How is performance presented to the user/how is degradation flagged
- How easy is end-user use and interface with the system

Train and Validate a model:

Achieve a nominal, acceptable level of performance (>95% e.g.)

Problems to address

- Imbalanced dataset, model bias
- Tuning
- Automated/User-friendly data pipeline
- Clear performance metrics

A trained and validated model in a .ipynb notebook

Assess model under test conditions

Possible issues that arise:

- Unacceptable performance degradation
- Incongruous data
- The need to retrain the model
- Contested/degraded operations

Enter a "new environment":

Ingest test data and assess model

Test Data

- 138 Small or Unclassifiable Ships
- 24 Blurred or Fractured Planes
- 199 New Empty Seabed/Seafloor images

A folder with 3 labeled subfolders each with images as stated

Rojo and Wreck stop here