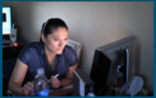
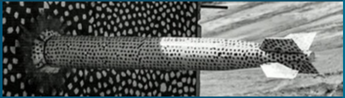
Seacape: A Due-Diligence Framework For Algorithm Acquisition





Presented by:

Christopher Pitts

Authors:

Christopher Pitts, Forest Danford, Emily Moore, William Marchetto, Henry Qiu, Leon Ross, Todd Pitts

SPIE Artificial Intelligence and Machine Learning in Defense Applications IV
September 7, 2022

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525.
SAND NO. SAND2022-11758C

Program-Level Problems in Algorithm Acquisition

A program tasked with algorithm acquisition (as opposed to development) faces some important problems to solve:

1. Management of sensor data representative of the needs of the program and the expected conditions of the algorithms in deployment
2. Provisioning of a means of auditing and validating algorithms developed by vendors, whether in-house or externally
3. Generation of assessments and reports on the performance of algorithms

Sensor Data Management - Data Types

The diverse nature of sensing equipment and platforms means a wide array of data types must be supported:

- Raw sensor values
- Expert-adjudicated labels
- Metadata that may be provided by a sensing platform

Sensor Data Management - Data Formats

Different sensing needs will require different data formats, which must also be supported by a program seeking to acquire and deploy algorithms.

- Image data
  - JPEG
  - TIFF
  - PNG
- Time series data
  - HDF5
  - CSV or raw
  - OGG
  - MP3
- Other data types and bespoke proprietary formats

Sensor Data Management - NRE

Non-recurring engineering (NRE) costs, specifically expert adjudication of labels for sensor data, can be quite high, particularly in niche or highly specialized domains. Ideally, they really are *non*-recurring, and work products from this effort can be appropriately stored for later use.

This point may be somewhat tautological, but it can be hard to quantify the true cost of having data that is not stored in a machine-readable, well-supported, and thoroughly documented format.

Seascape Data Management

Seascape defines a program-level storage interface for data. From the perspective of an algorithm or a report generation plugin, the exact nature of underlying storage is hidden behind an abstraction layer that defines the minimal set of operations necessary to retrieve data from storage.

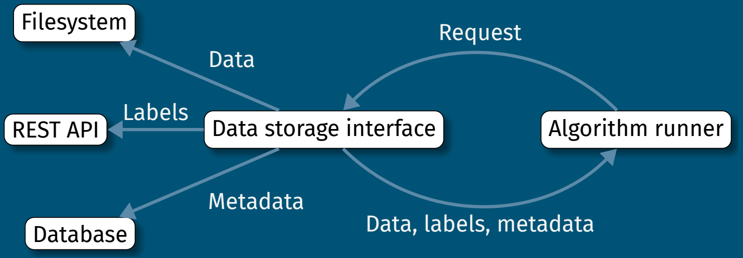


Figure 1: Storage abstractions simplify interactions

Algorithm Validation - Desired Characteristics

An algorithm validation process should possess certain characteristics:

1. Automated
2. Reproducible
3. Easily analyzable
4. Scalable

These principles are often invoked as part of Machine Learning Operations (MLOps), although we consider them to be principles of due diligence in evaluating *any* kind of algorithm.

Algorithm Validation - Automation

- Manual validation processes requiring experts are expensive, prone to error, and lead to significant program risk of omissions, errors, and lack of repeatability
- Automated analysis of algorithms significantly lowers the cost of testing. It also assures that every version of an algorithm is tested, every time it is changed
- Seascape defines an automated validation process that pulls data and test definitions from storage, executes an algorithm, and outputs results

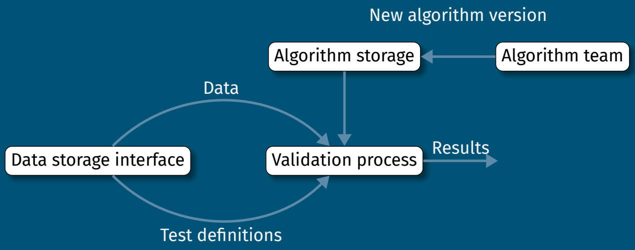


Figure 2: Automated validation - every change, every algorithm

Algorithm Validation - Reproducibility

- Reproducible results are a cornerstone of responsible science and engineering
- A test you can't reproduce isn't really a test
- Seascape implements algorithm storage in Git and leverages commit hashes to directly tie specific versions of algorithms to specific results

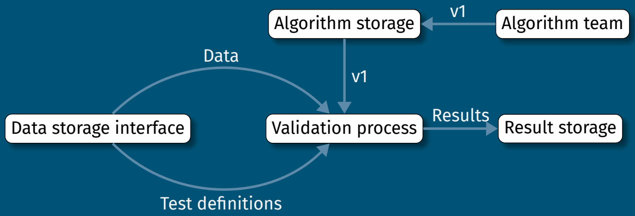


Figure 3: Version control hashes tie algorithm versions to results

Algorithm Validation - Reproducibility

Being able to directly associate algorithm versions with results enables a number of other beneficial capabilities:

- Longitudinal performance studies - are we getting better or worse?
- Breaking change detection - what changes were tied to a drop in performance?

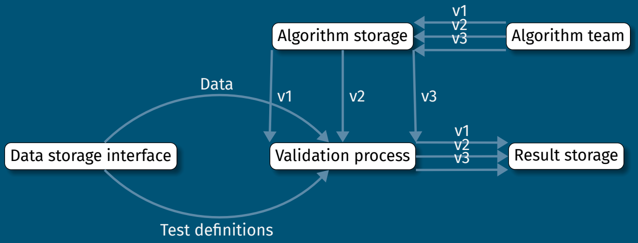


Figure 4: Using hashes as keys enables 1-to-1 comparisons

Algorithm Validation - Analysis

- Tabulation and scoring of results, including generation of metrics and figures to aid understanding, is a key need of algorithm acquisition
- The choice of metrics is an important factor in evaluating an algorithm
- Seascape enables analysis of results by using program-defined metric computation and report generation, ensuring standardized comparisons across diverse algorithms and vendors

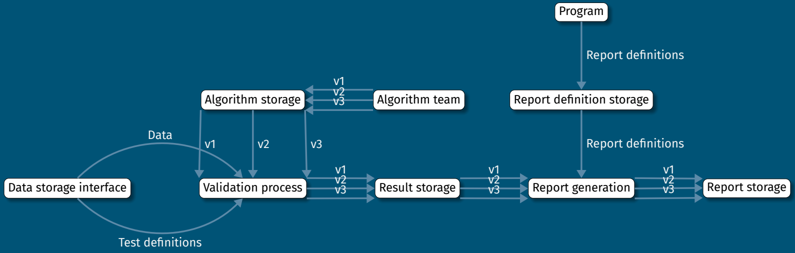


Figure 5: Programs can specify uniform reports across vendors

Algorithm Validation - Scalability

- Efficient use of computational resources is critical for success in algorithm acquisition
- Constraints on network bandwidth, hardware resources such as GPUs, RAM, and physical cores, and the number of computing nodes can limit the ability to assess certain algorithms
- Seascape enables scalable computing by batching images to dispatch to algorithm runners, powered by GitLab CI and GitLab Runner

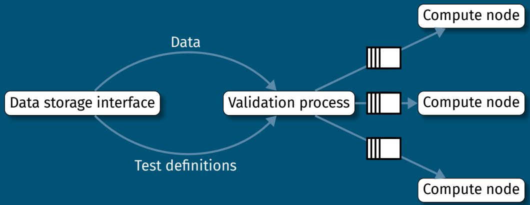
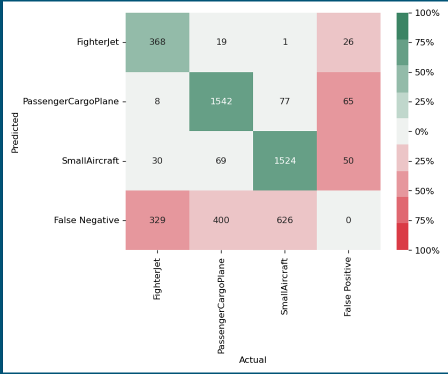


Figure 6: Batch dispatch prevents resource hogging

Empirical Results - Seascape Report Artifacts

(a)



(b)

Figure 7: Confusion matrix plots for RetinaNet (a) and DetectNet (b)

Empirical Results - Seascape Report Artifacts

Algorithm name (hash) Test Date	**retinanet (6e84)** 2022-03-25	**detectnet (94d1)** 2022-03-28
FighterJet (735)		
True positive	368	525
False negative	367	210
False positive	46	54
Precision	0.50	0.71
Recall	0.89	0.91
F1	0.64	0.80
PassengerCargoPlane (2029)		
True positive	1542	1873
False negative	488	156
False positive	150	1049
Precision	0.76	0.92
Recall	0.91	0.64
F1	0.83	0.76
SmallAircraft (2228)		
True positive	1524	706
False negative	704	1527
False positive	149	101
Precision	0.68	0.32
Recall	0.91	0.87
F1	0.78	0.46

Table 1: Numerical result data for overhead plane imagery

Acknowledgments

This work was conducted with funding from the Navy RSCD (Remote Sensing Capability Development) program operating in PMW-120 (Battlespace Awareness and Information Operations Program Office) under PEO C4I (Program Executive Office Command, Control, Communications, Computers, Intelligence and Space Systems), contract number N0003922IP00013. Some Seascape capabilities were developed with funding from the Air Force Air Combat Command under contract number F3QCAZ0161G101. Approved for unclassified unlimited release (UUR) by Sandia National Laboratories, SAND Number SAND2022-11758 C.