Carnegie Mellon University
Software Engineering Institute

Flexible and Extensible Drift Monitoring in Machine Learning Systems

FEBRUARY 25, 2025

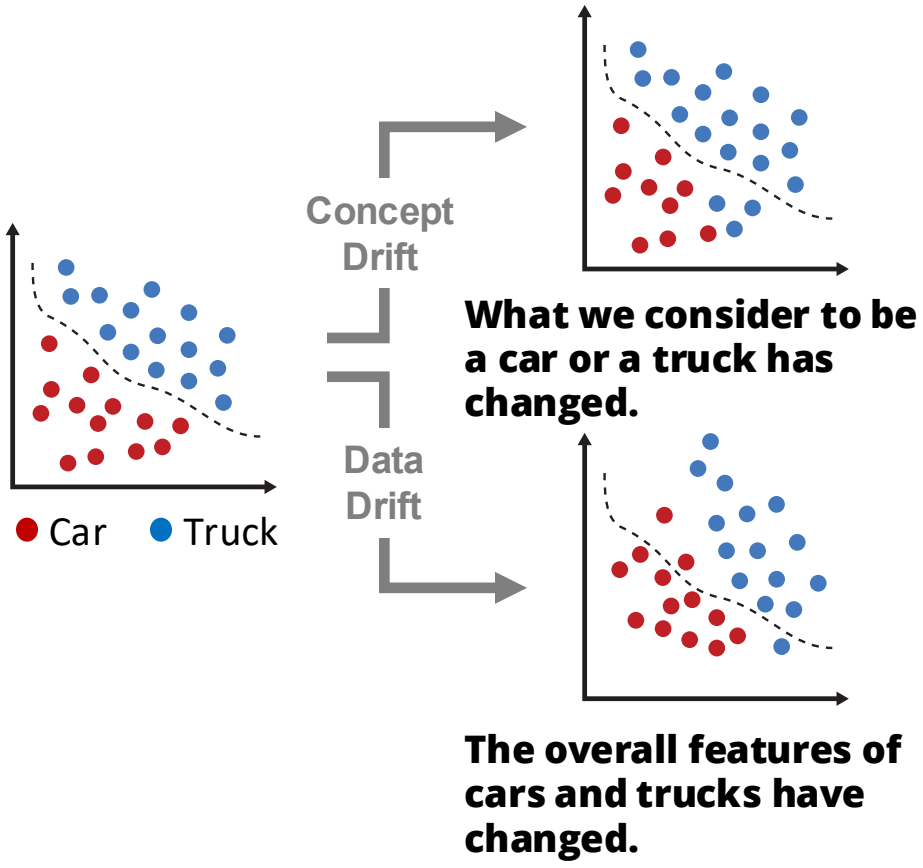
Ipek Ozkaya, ozkaya@sei.cmu.edu
Technical Director, Engineering Intelligent Software Systems

Document Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM25-0101

Drift Detection and Management

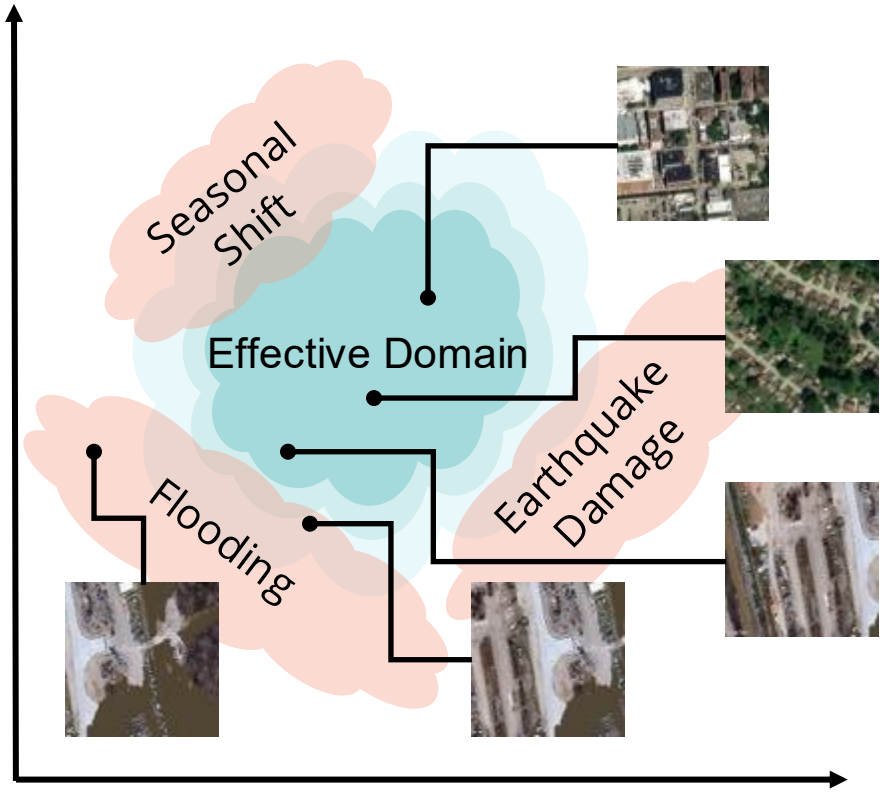


The performance of an ML-enabled system declines over time due to changes in the data environment

Drift can result in silent failures of the model that impact mission outcomes and **negatively impact operator trust**

There are several approaches to restoring system performance including: periodic retraining, online learning, adversarial model training and **drift planning**

Drift Monitoring Provides Active Feedback



ML models are powerful but brittle

- **Unexpected inputs** can lead to undesirable behavior
- Systems often **do not know** when they are performing poorly leading to mission failure

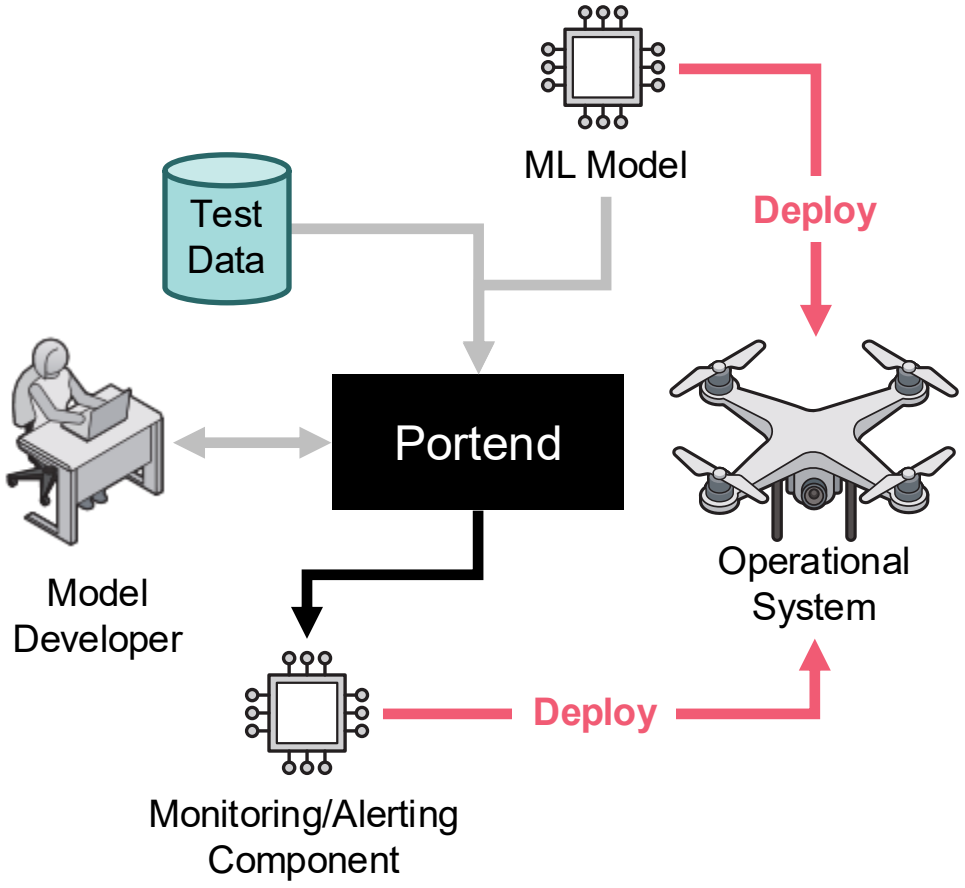
Drift monitoring using Portend enables active feedback by

- providing **near-real time** estimates of model performance on operational data
- enable systems to **take alternative actions** or **schedule retraining** when model performance is degraded

Data Drift and ML-enabled Systems

Carnegie Mellon University
Software Engineering Institute

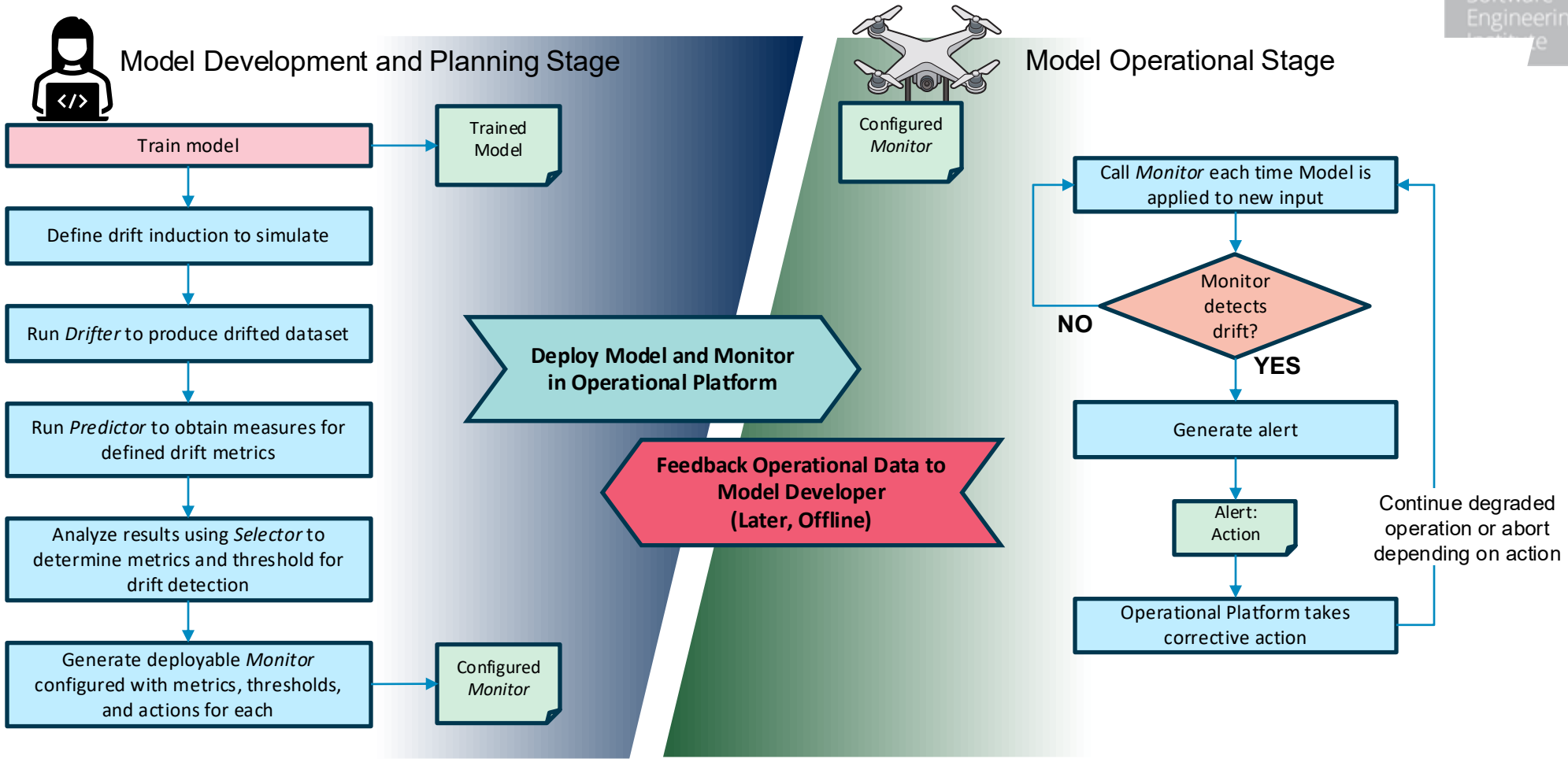
Drift Planning Enables Better Human-Machine Teaming



The Portend tool set can be used to the following drift planning process:

- Simulate model performance degradation under controlled drift conditions
- Collect simulation data for one or more drift detection metrics
- Use the results of the simulation to select appropriate drift detection metrics and set thresholds at which to trigger some downstream actions (e.g. alerts, take the system offline, bring human into the loop)

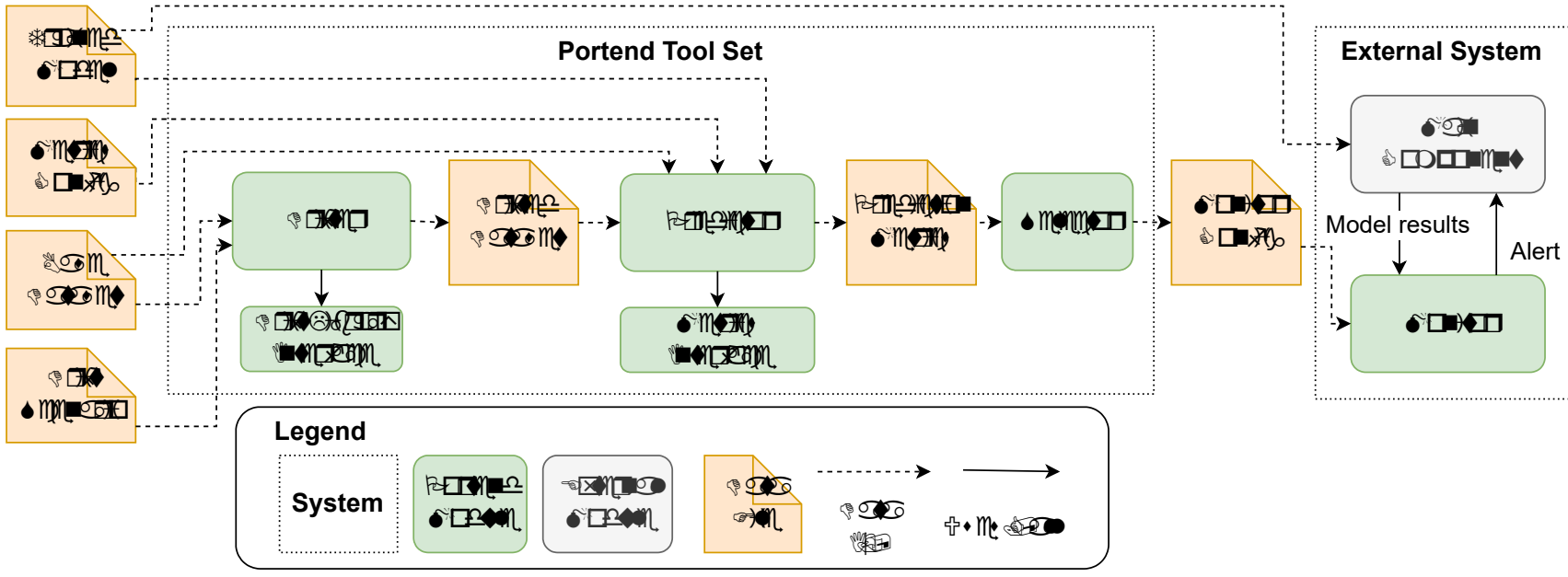
Portend Workflows



Portend Tool Set for Drift Planning

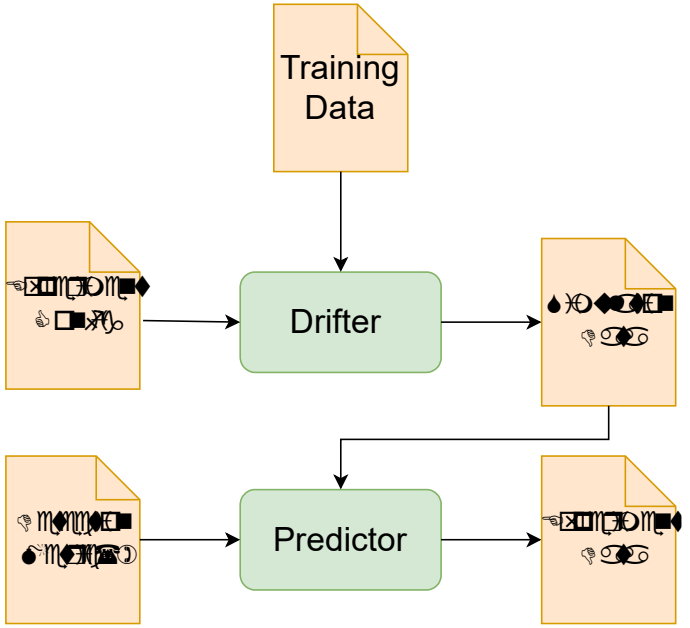
Carnegie Mellon University
Software Engineering Institute

Portend Tool Set Overview

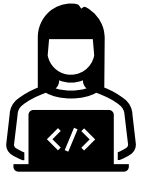


Portend Tool Set – Planning Stage

Portend is extensible. It can take a configuration for “any” model and data to simulate drift for any data type if the drift induction method has been implemented.

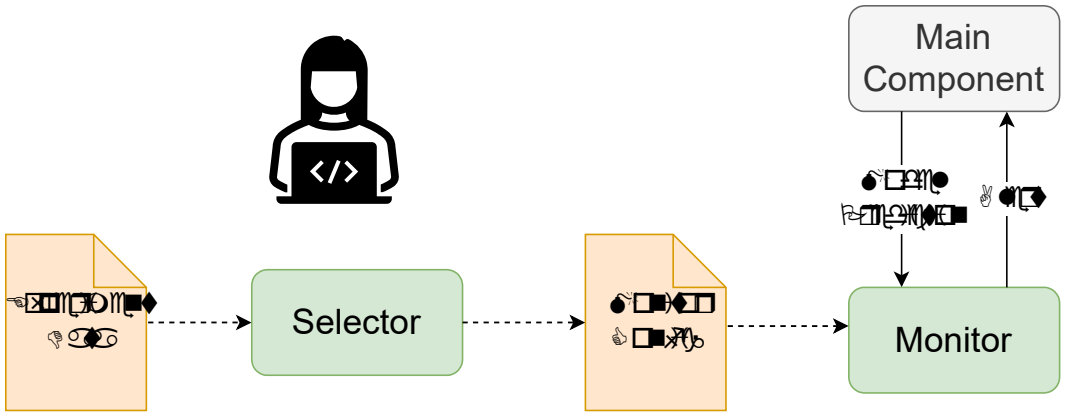


A model developer can then review the experiment data and use Portend tools to select one or more monitors that can be deployed to the operational model.



Portend Tool Set – Monitor Tuning

Model developer reviews experiment data and uses built in functions in the *Selector* to select drift detection metrics and set performance thresholds.



The *Monitor* includes the drift detection metrics and thresholds to take corrective actions, e.g. firing an alert.

Demonstration in a Visual Localization System

Carnegie Mellon University
Software Engineering Institute

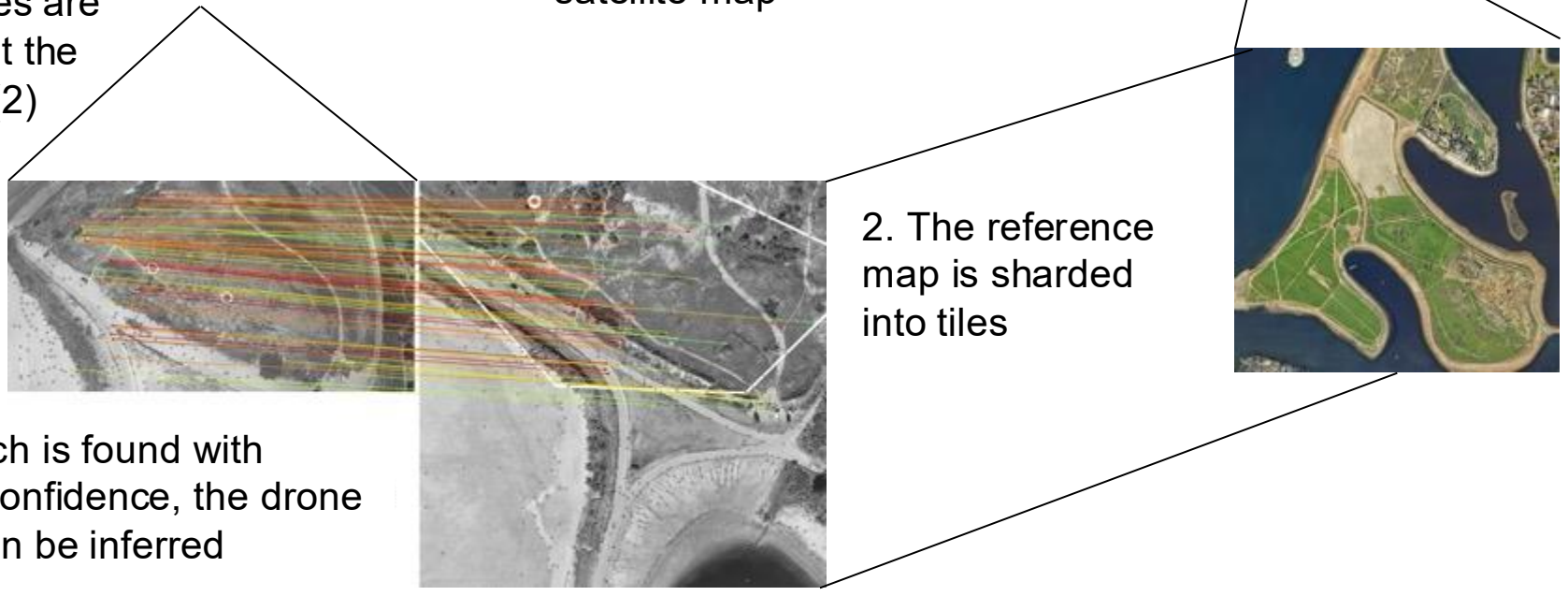
Visual Localization

1. A CNN model is trained on a reference region from a satellite map

2. The reference map is sharded into tiles

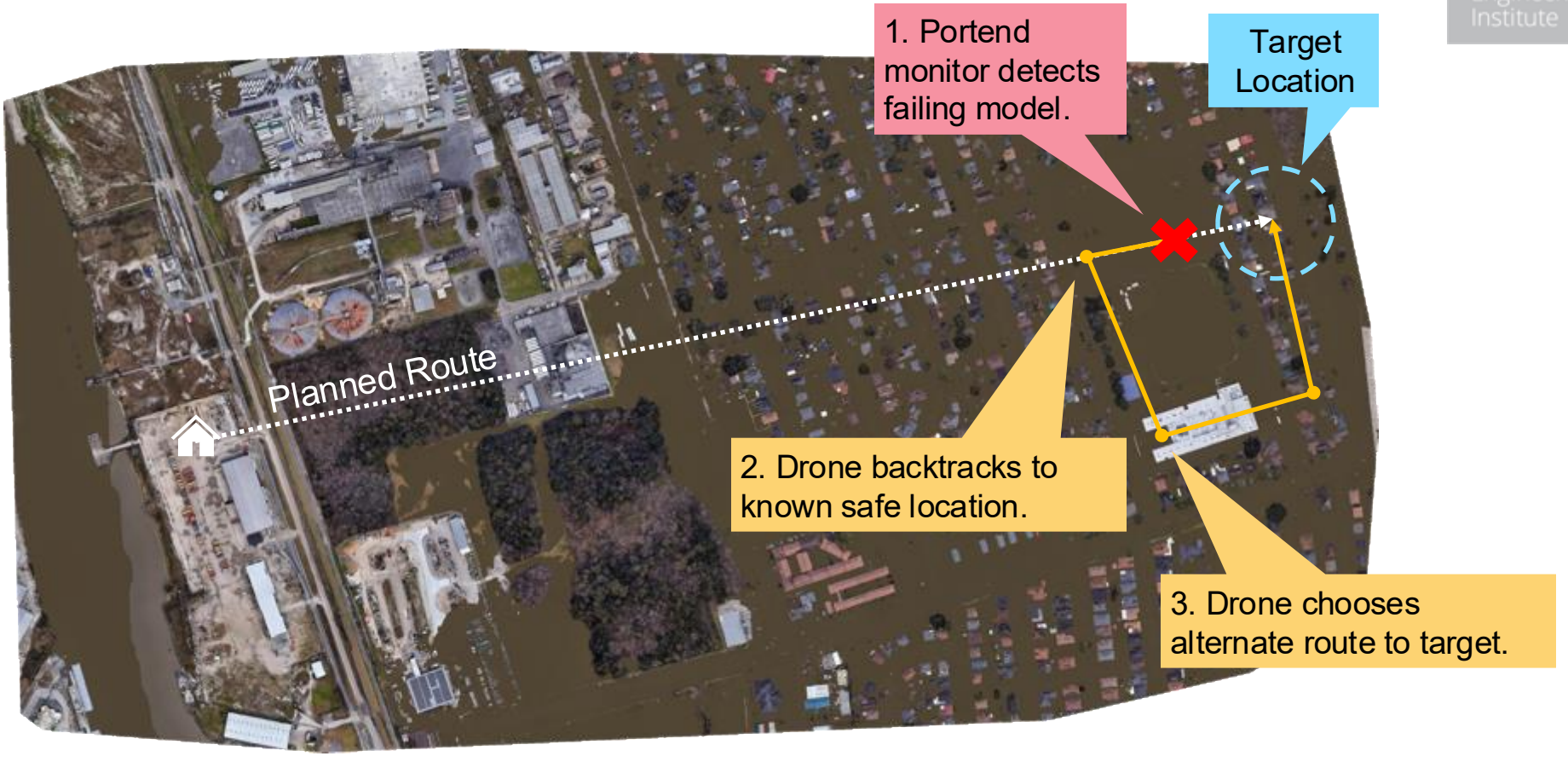
3. Points extracted* from UAV images are matched against the reference map (2)

4. If a match is found with sufficient confidence, the drone position can be inferred



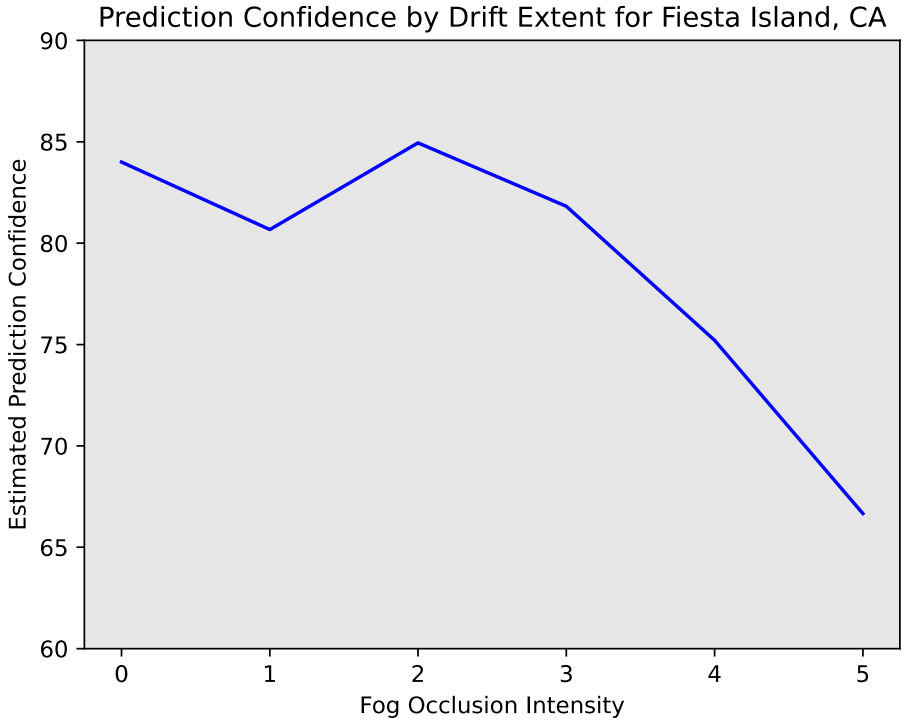
*M.-M. Gurgu, J. P. Queralta, and T. Westerlund, “Vision-Based GNSSFree Localization for UAVs in the Wild,” arXiv, 2022.

Portend-Enabled Autonomous Navigation



Drift Planning for Visual Localization

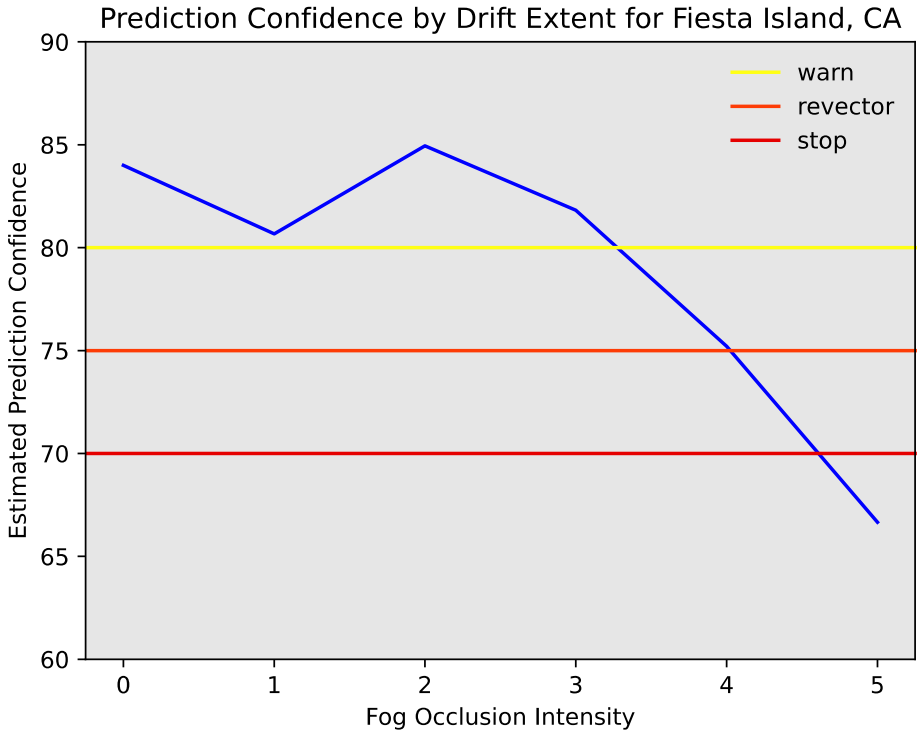
- We simulated fog drift for five degrees of increasing occlusion for a dataset of 225 overhead satellite images over Fiesta Island, CA.
- We computed a drift metric based on the average threshold confidence* to estimate the prediction confidence



* S. Garg, S. et al, "Leveraging Unlabeled Data to Predict Out-of-Distribution Performance," CoRR, vol. abs/2201.04234, 2022. [Online]. Available: https://arxiv.org/abs/2201.04234

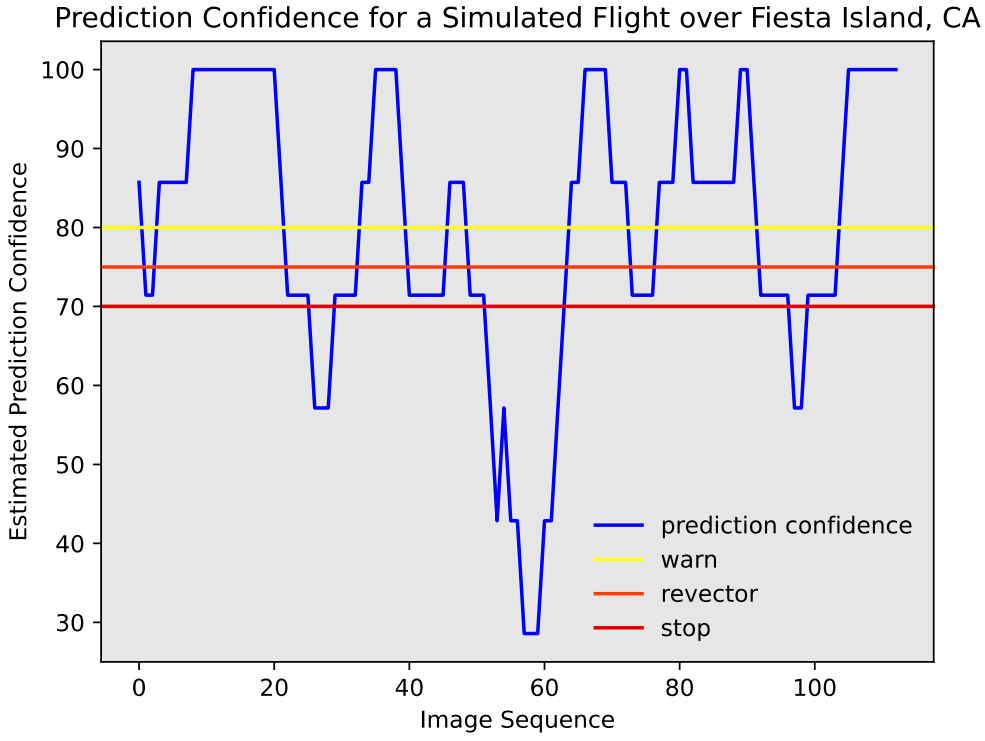
Threshold Estimation for Fog Drifted Images

- Threshold setting requires input from the domain needs as well as the response of the metric to drift.
- We selected thresholds based on the rate of performance degradation as drift increases.
- In some applications, there may be a minimum acceptable performance threshold.



Monitoring Drift in a Simulated Flight

- These data show the prediction confidence over a sequence of images that approximate flight over a contiguous region.
- The prediction confidence is computed over a sliding window to give an operator an idea of the model confidence for a particular region.
- In regions where the estimated prediction confidence is below one of the defined thresholds, an alert is logged.



How to Use Portend Planning Tools

- Portend currently supports
  - Drift libraries to simulate time series effects (4 conditions) and image manipulation for fog and flood
  - 7 distribution test drift detection metrics and 1 error based metric (ATC)
  - Overhead image data
  - Producing a monitor that directly integrates as python code
- Portend can be extended to support additional planning use cases with
  - Trained model, data, and problem context
  - Monitors can be ported to other programming languages

Open Source Portend tools: https://github.com/cmu-sei/portend

Portend Team

Dr. Jeffrey Hansen
Portend PI, Senior Machine Learning Research Scientist
Email: jhansen@sei.cmu.edu

Sebastian Echeverria
Senior Software Engineer

Lena Pons
Machine Learning Scientist

Lihan Zhan
Assistant Software Engineer

Dr. Gabriel Moreno
Principal Researcher

Dr. Grace Lewis
Technical Advisor
Principal Researcher and Tactical and AI-Enabled Systems Initiative Lead

Dr. Ipek Ozkaya
Technical Advisor
Technical Director
Engineering Intelligent Software Systems