Efficient out-of-distribution detection for reliable deployment of DNNs

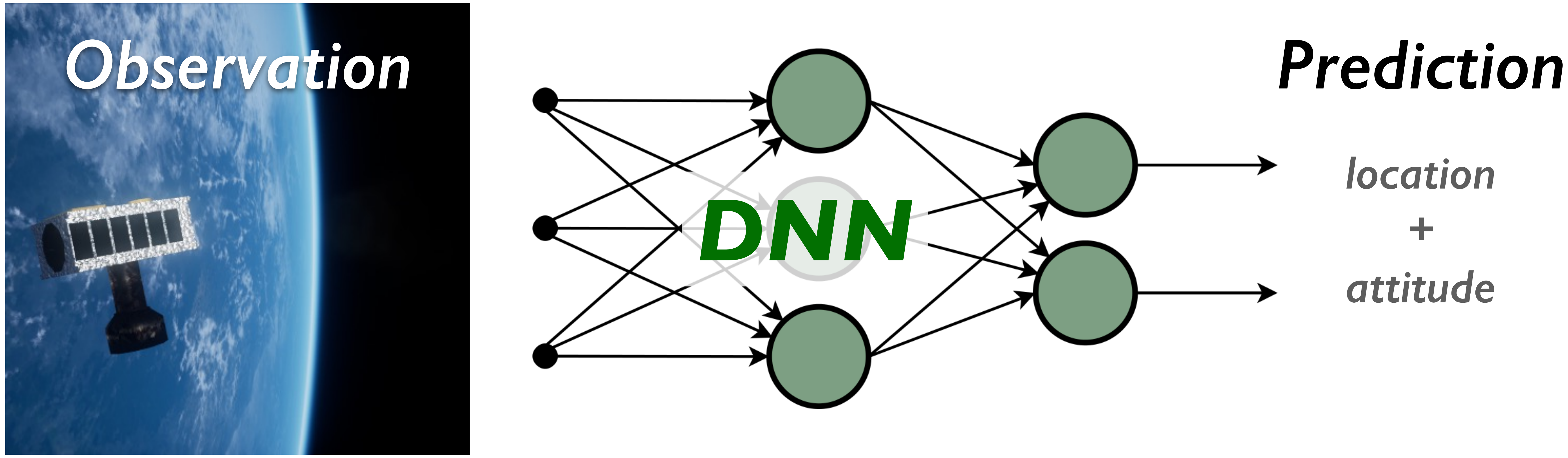
Apoorva Sharma
joint work with Somrita Banerjee, Navid Azizan, and Marco Pavone

GSAW 2022

Approved for public release

Copyright by the authors. Published by The Aerospace Corporation with permission.

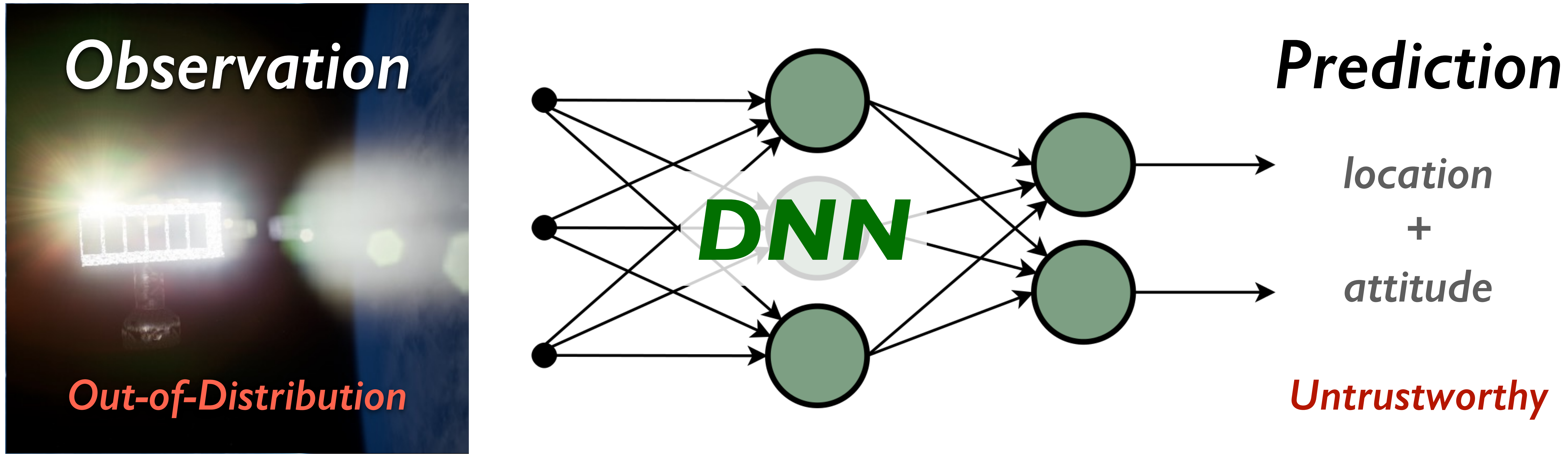
Machine learning tools can provide key capabilities for space ground systems



DNNs can provide **data-driven predictions** in **real-time** on **high-dimensional** perceptual inputs

Copyright by the authors. Published by The Aerospace Corporation with permission.

Machine learning tools can provide key capabilities for space ground systems

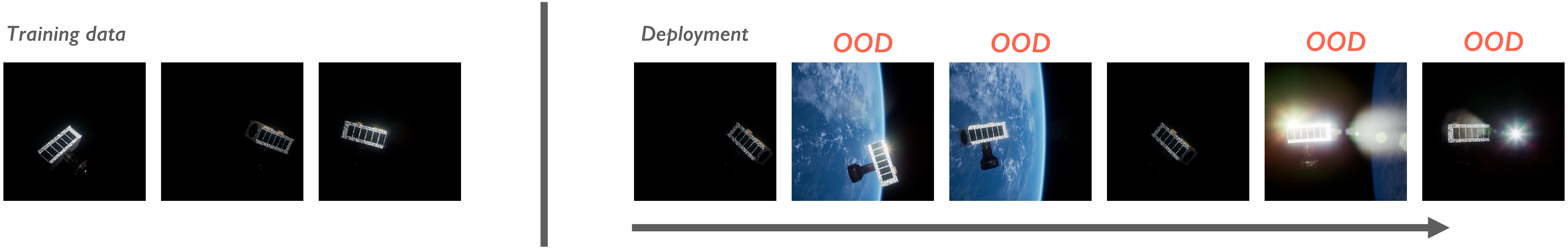


DNNs can provide **data-driven predictions** in **real-time** on **high-dimensional** perceptual inputs

However, they can suffer from **poor reliability** in conditions that **deviate from training data.**

Copyright by the authors. Published by The Aerospace Corporation with permission.

Ensuring reliable operation of DNNs requires detecting and reacting to changing conditions.



How can we efficiently detect anomalous conditions during operation?

How can we efficiently retrain DNN models to adapt to changing conditions?

Copyright by the authors. Published by The Aerospace Corporation with permission.

Outline

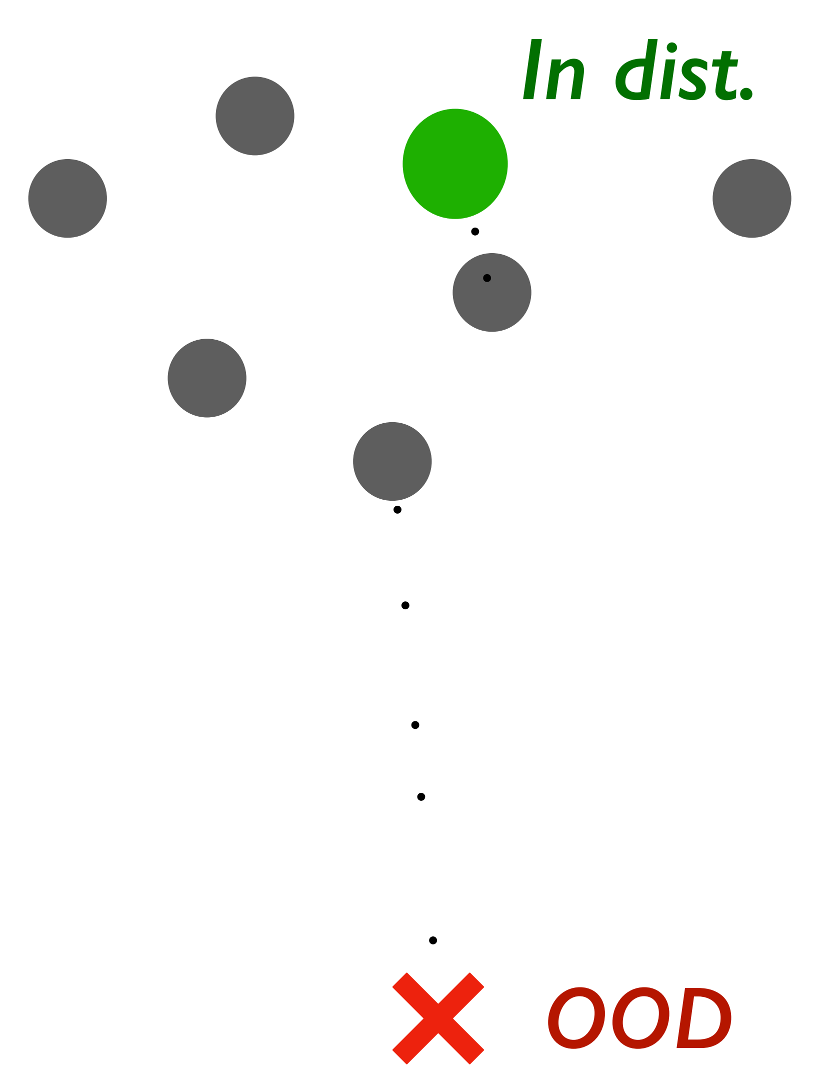
- Introduction
- **Efficient anomaly detection for pre-trained DNNs**
  - Problem Setup
  - SCOD: Sketching Curvature for Out-of-Distribution Detection
  - Insights and Results
- Leveraging out-of-distribution detection in the data-collection pipeline

Copyright by the authors. Published by The Aerospace Corporation with permission.

Defining “out-of-distribution”

Distance based:

How far away is a new data point to training data?



Intuitive, easy to implement

What distance metric to use?

Need to hold on to training data at test time

Copyright by the authors. Published by The Aerospace Corporation with permission.

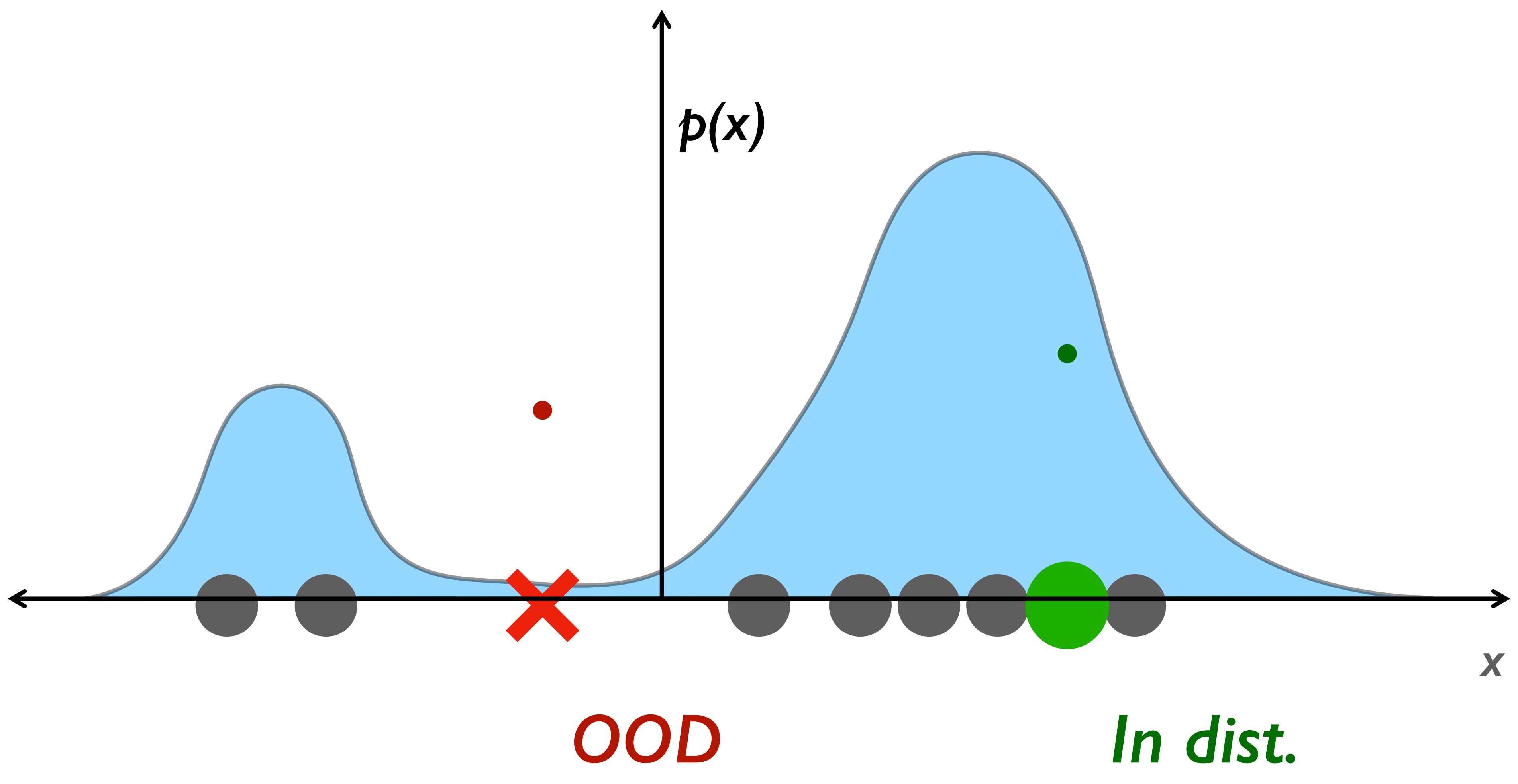
Defining "out-of-distribution"

Distance based:

How far away is a new data point to training data?

Distribution based:

Can we compare test-time data against the training data distribution?



Parametric distribution can summarize large dataset

How to model distribution over high-dimensional inputs?
How to evaluate correlated test-time inputs?

Copyright by the authors. Published by The Aerospace Corporation with permission.

Defining "out-of-distribution"

Distance based:

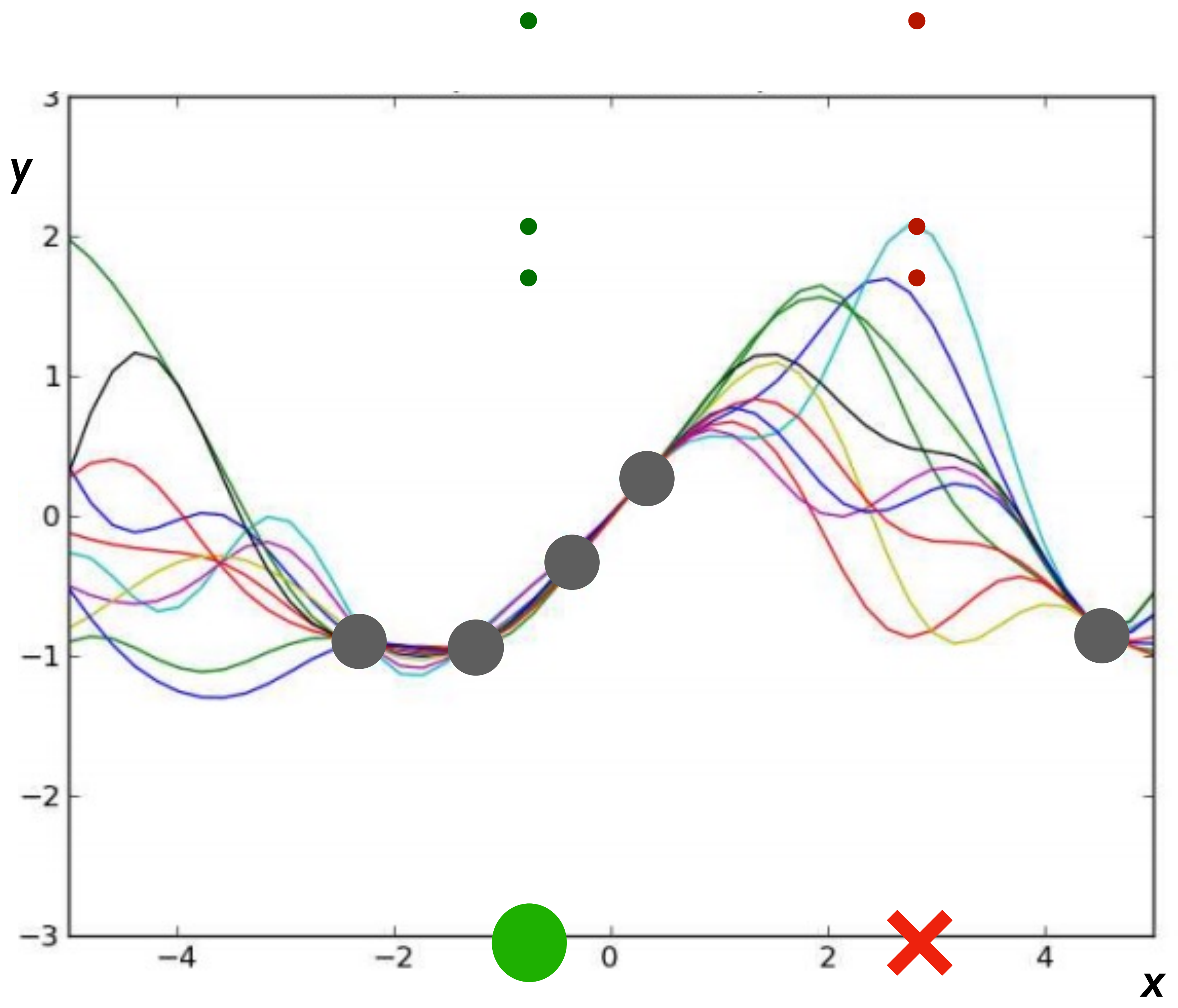
How far away is a new data point to training data?

Distribution based:

Can we compare test-time data against the training data distribution?

Functional uncertainty:

What outputs are still likely for a test-time input given the training data?



Accounts for input-output relationship
Useful for reasoning about adaptation

How to quantify functional uncertainty?

Copyright by the authors. Published by The Aerospace Corporation with permission.

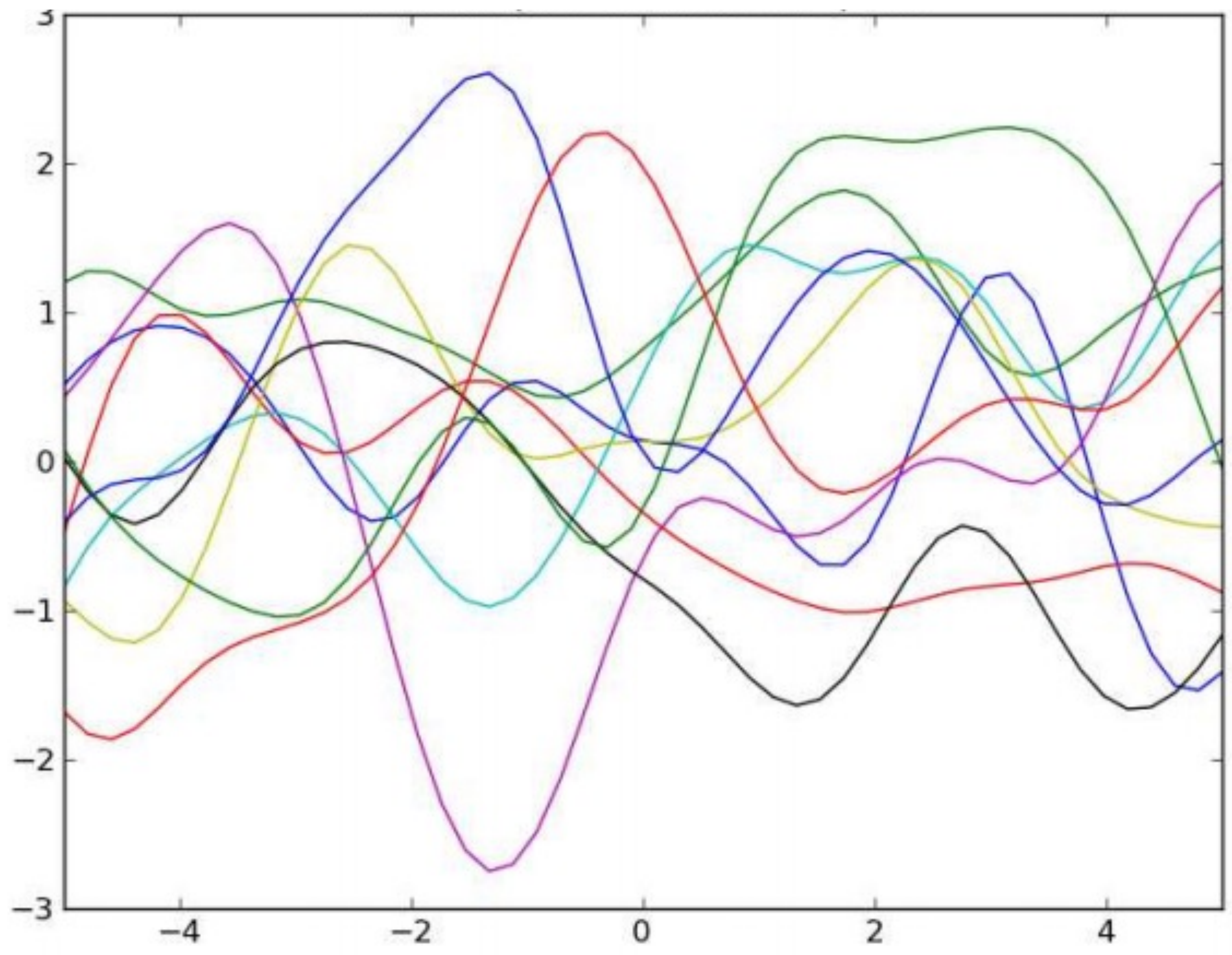
Bayesian methods offer a principled approach to quantifying functional uncertainty

Basic formula:

1. **Propose a broad prior over the space of functions mapping inputs to outputs.**

2. Given training data, compute posterior in function space.

3. Treat inputs with high posterior uncertainty as anomalous.



Copyright by the authors. Published by The Aerospace Corporation with permission.

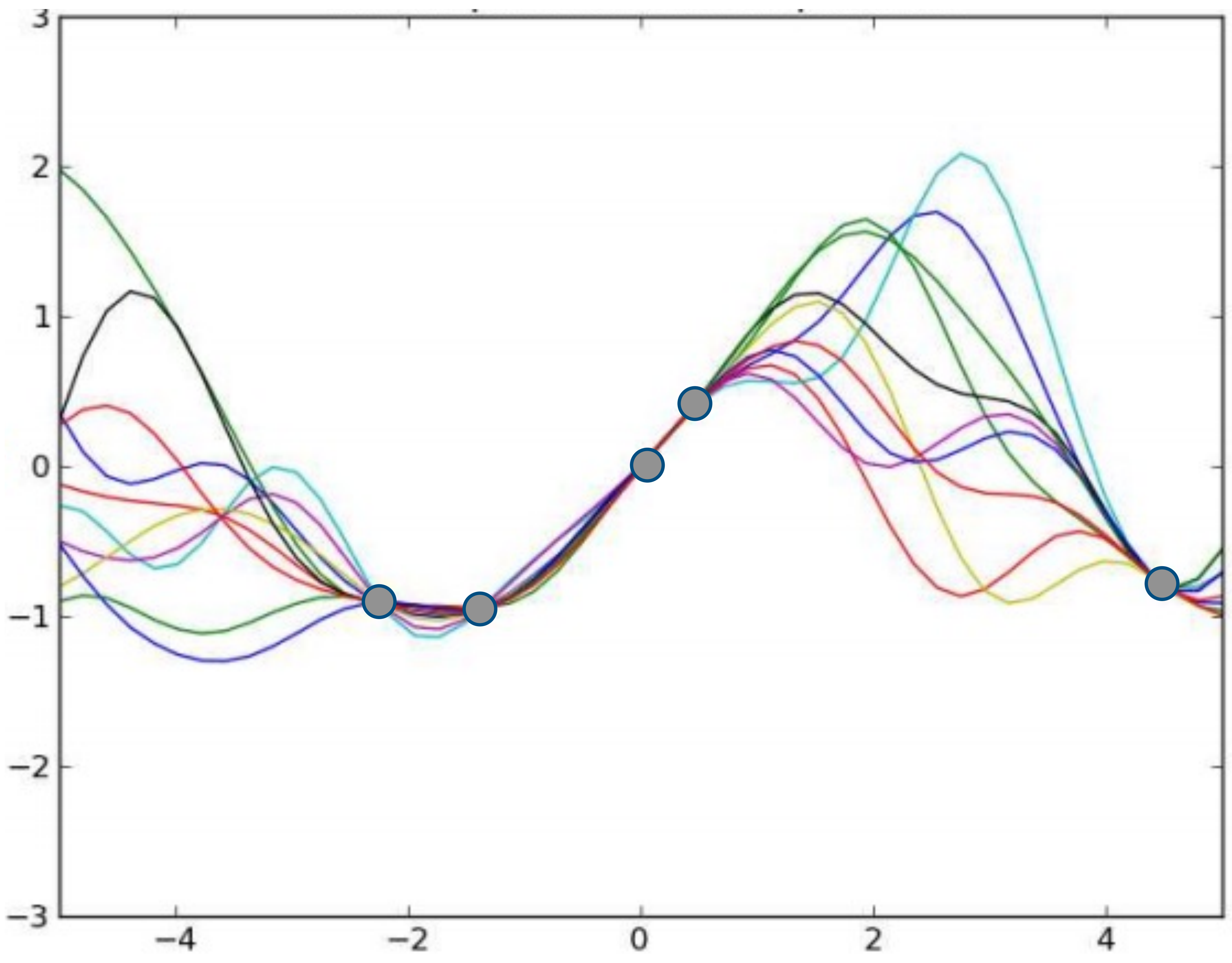
Bayesian methods offer a principled approach to quantifying functional uncertainty

Basic formula:

1. Propose a broad prior over the space of functions mapping inputs to outputs.

2. **Given training data, compute posterior in function space.**

3. Treat inputs with high posterior uncertainty as anomalous.



Copyright by the authors. Published by The Aerospace Corporation with permission.

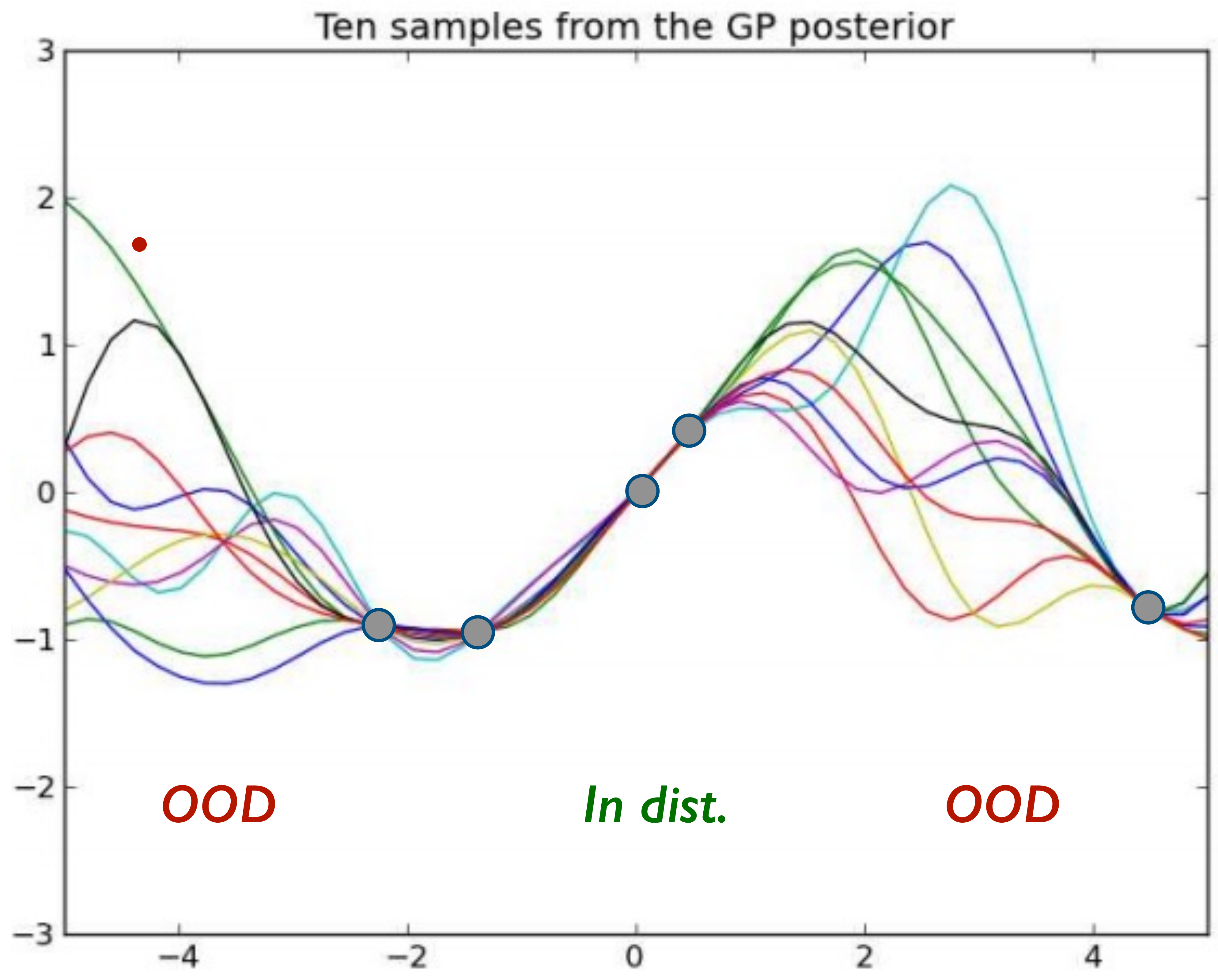
Bayesian methods offer a principled approach to quantifying functional uncertainty

Basic formula:

1. Propose a broad prior over the space of functions mapping inputs to outputs.

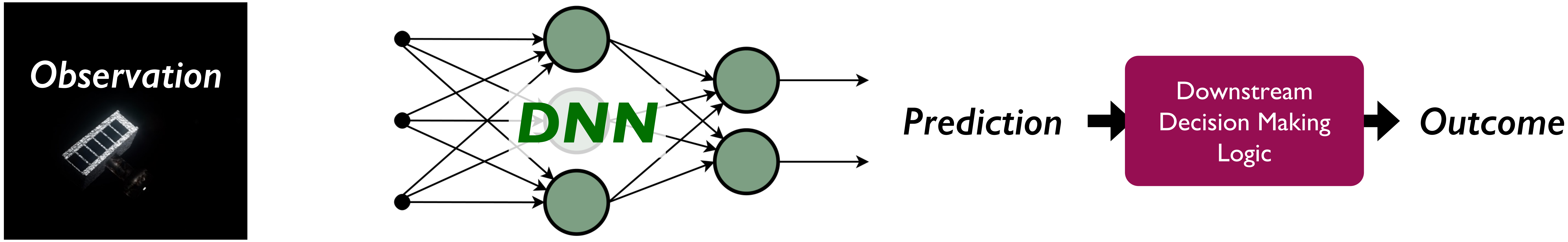
2. Given training data, compute posterior in function space.

3. **Treat inputs with high posterior uncertainty as anomalous.**



Copyright by the authors. Published by The Aerospace Corporation with permission.

How can we reason about functional uncertainty for real-time anomaly detection?



Good functional prior

Need a task-aligned prior over functions on high-dimensional sensor input

Efficient posterior estimation and representation

Want a memory-efficient posterior representation which summarizes the training data

Efficient predictive uncertainty computation

Need to compute functional uncertainty at test inputs with low latency

Copyright by the authors. Published by The Aerospace Corporation with permission.

Outline

- Introduction
- Efficient anomaly detection for pre-trained DNNs
  - Problem Setup
  - **SCOD: Sketching Curvature for Out-of-Distribution Detection**
  - Insights and Results
- Leveraging out-of-distribution detection in the data-collection pipeline

Copyright by the authors. Published by The Aerospace Corporation with permission.

SCOD: Sketching Curvature for OOD detection

SCOD addresses these requirements through careful design decisions

Good functional prior

Leverage existing task DNN to create parametric prior

Efficient posterior estimation and representation

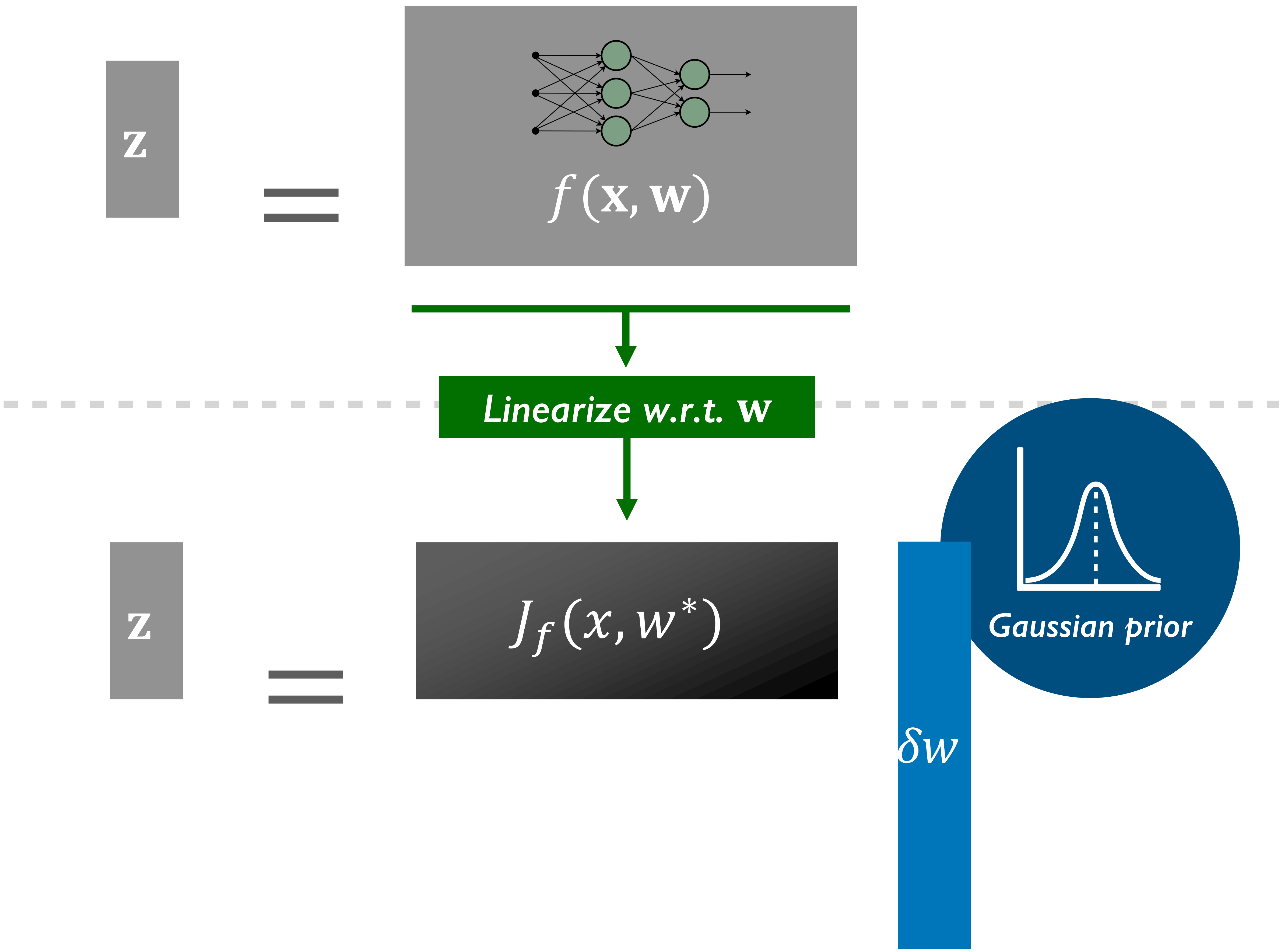
Low-rank posterior representation via matrix sketching

Efficient predictive uncertainty computation

Sampling-free predictive uncertainty computation

Copyright by the authors. Published by The Aerospace Corporation with permission.

SCOD quantifies uncertainty in a DNN by applying Bayesian analysis to a surrogate linear model.



Wide and aligned prior:

Leverages task-specific structure of existing, pre-trained DNN

Tractable posterior computation:

Low-rank approximation via matrix sketching mitigates memory bottlenecks

Efficient predictive uncertainty estimation:

Linearized model allows for direct posterior predictive uncertainty computation, without Monte-Carlo sampling

Copyright by the authors. Published by The Aerospace Corporation with permission.

SCOD: Sketching Curvature for OoD Detection

Algorithm Overview

DNN with optimized weights

$$\mathbf{z} = f(\mathbf{x}, \mathbf{w}^*)$$

Output distribution (e.g. Gaussian, Categorical)

$$p(\mathbf{y} \mid \mathbf{z})$$

Training Dataset

$$\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{M}$$

Linearize model

$$\mathbf{z} \approx f(\mathbf{x}, \mathbf{w}^*) + J_f(\mathbf{x}, \mathbf{w}^*) \cdot \delta\mathbf{w}$$

Offline

Compute Posterior Distribution on $\delta\mathbf{w}$

$$p(\delta\mathbf{w} \mid \mathcal{D}) = \mathcal{N}(\delta\mathbf{w}; \mathbf{0}, \Sigma_{\mathbf{w}})$$

Online

Compute posterior predictive distribution for linearized model

$$p(\mathbf{z} \mid \mathbf{x}, \mathcal{D}) = \mathcal{N}(\mathbf{z}; f(\mathbf{x}, \mathbf{w}^*), J_f(\mathbf{x}, \mathbf{w}^*) \Sigma_{\mathbf{w}} J_f(\mathbf{x}, \mathbf{w}^*)^\top)$$

and overall uncertainty score

$$\mathrm{Unc}(\mathbf{x} \mid \mathcal{D}) = \mathrm{Entropy}\left[\int p(\mathbf{y} \mid \mathbf{z}) p(\mathbf{z} \mid \mathbf{x}, \mathcal{D}) d\mathbf{z}\right]$$

Copyright by the authors. Published by The Aerospace Corporation with permission.

SCOD: Sketching Curvature for OoD Detection

Estimating and representing the posterior covariance $\Sigma_{\mathbf{w}}$

Analytic expression for posterior covariance involving only local curvature of DNN (Gauss Newton matrix)

$$\Sigma_{\mathbf{w}} = \left(\sigma_0^{-2} I + \sum_{i=1}^{M} J_f(\mathbf{x}^i, \mathbf{w}^*)^\top F_{\mathbf{z}}(\mathbf{z}^i) J_f(\mathbf{x}^i, \mathbf{w}^*)\right)^{-1}$$

Fisher information matrix of output distribution

Sketching based eigenvalue decomposition

Represent in terms of low-rank factors

$U \in \mathbb{R}^{N\times d}, \lambda \in \mathbb{R}^d$

$$\Sigma_{\mathbf{w}} \approx (\sigma_0^{-2} I + U \Lambda U^\top)^{-1}$$

Woodbury Matrix Identity

Never need to realize full NxN matrix

$$\Sigma_{\mathbf{w}} \approx \sigma_0^2 (I - U D U^\top)$$

Copyright by the authors. Published by The Aerospace Corporation with permission.

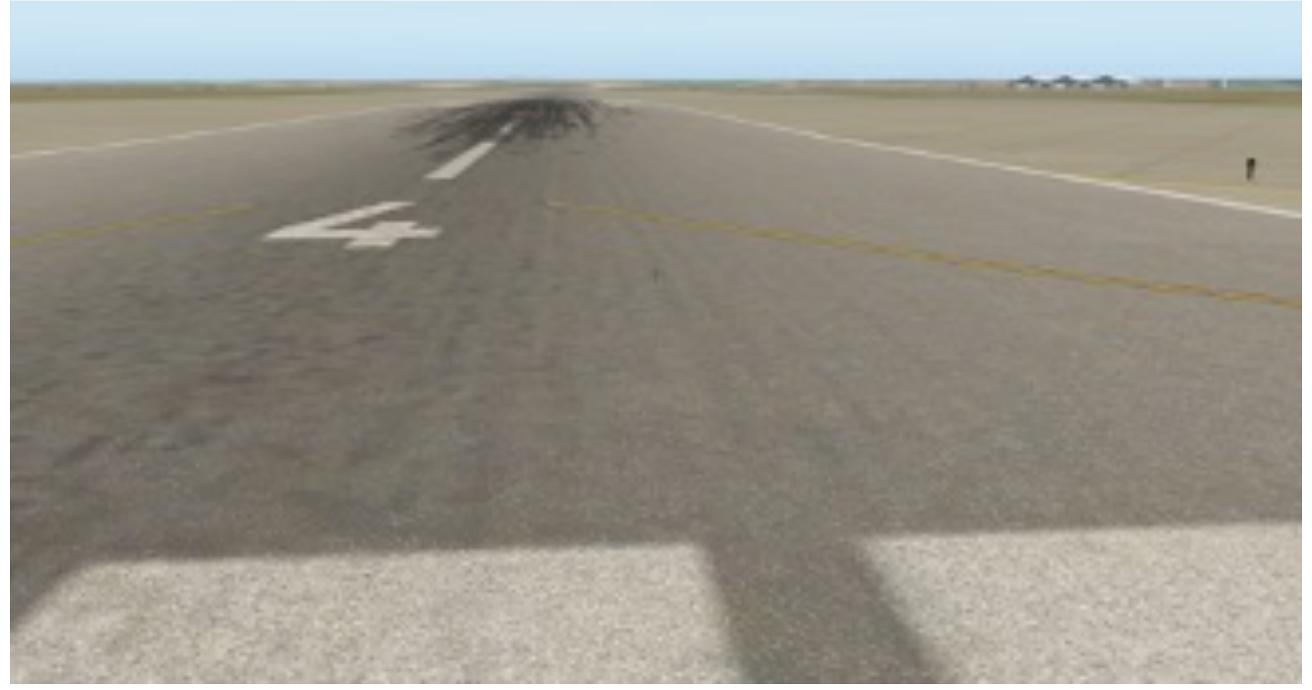
Outline

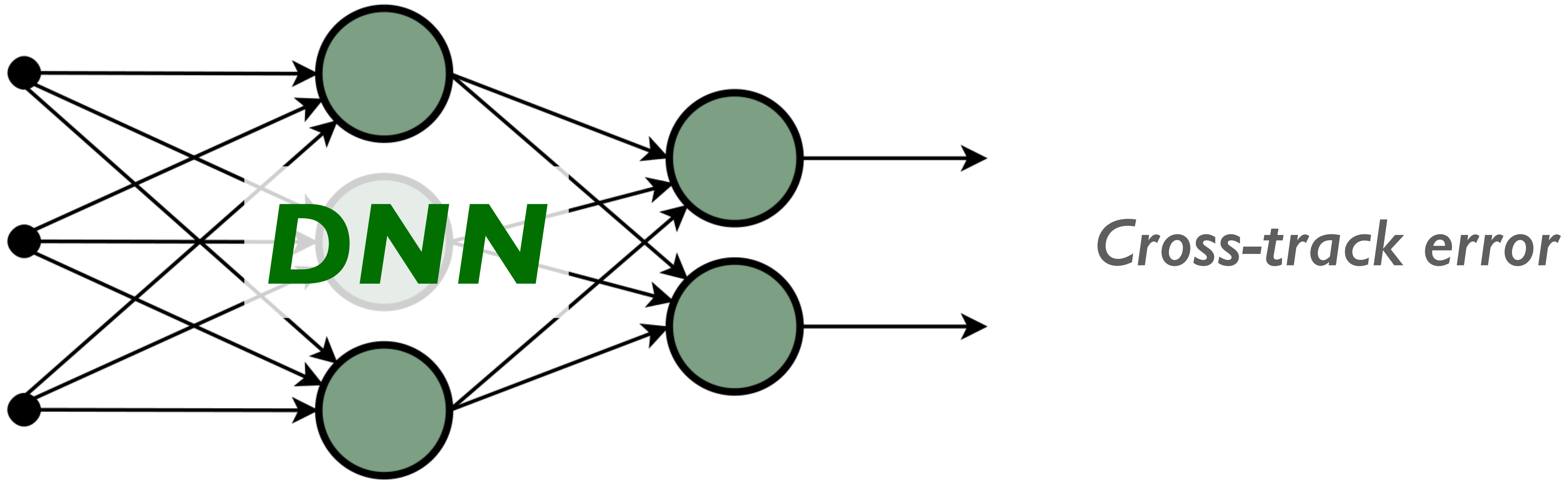
- Introduction
- Efficient anomaly detection for pre-trained DNNs
  - Problem Setup
  - SCOD: Sketching Curvature for Out-of-Distribution Detection
  - **Insights and Results**
- Leveraging out-of-distribution detection in the data-collection pipeline

Copyright by the authors. Published by The Aerospace Corporation with permission.

Qualitative Case Study

Visuomotor control of autonomous aircraft taxiing





Trained on simulated data from clear weather, early morning

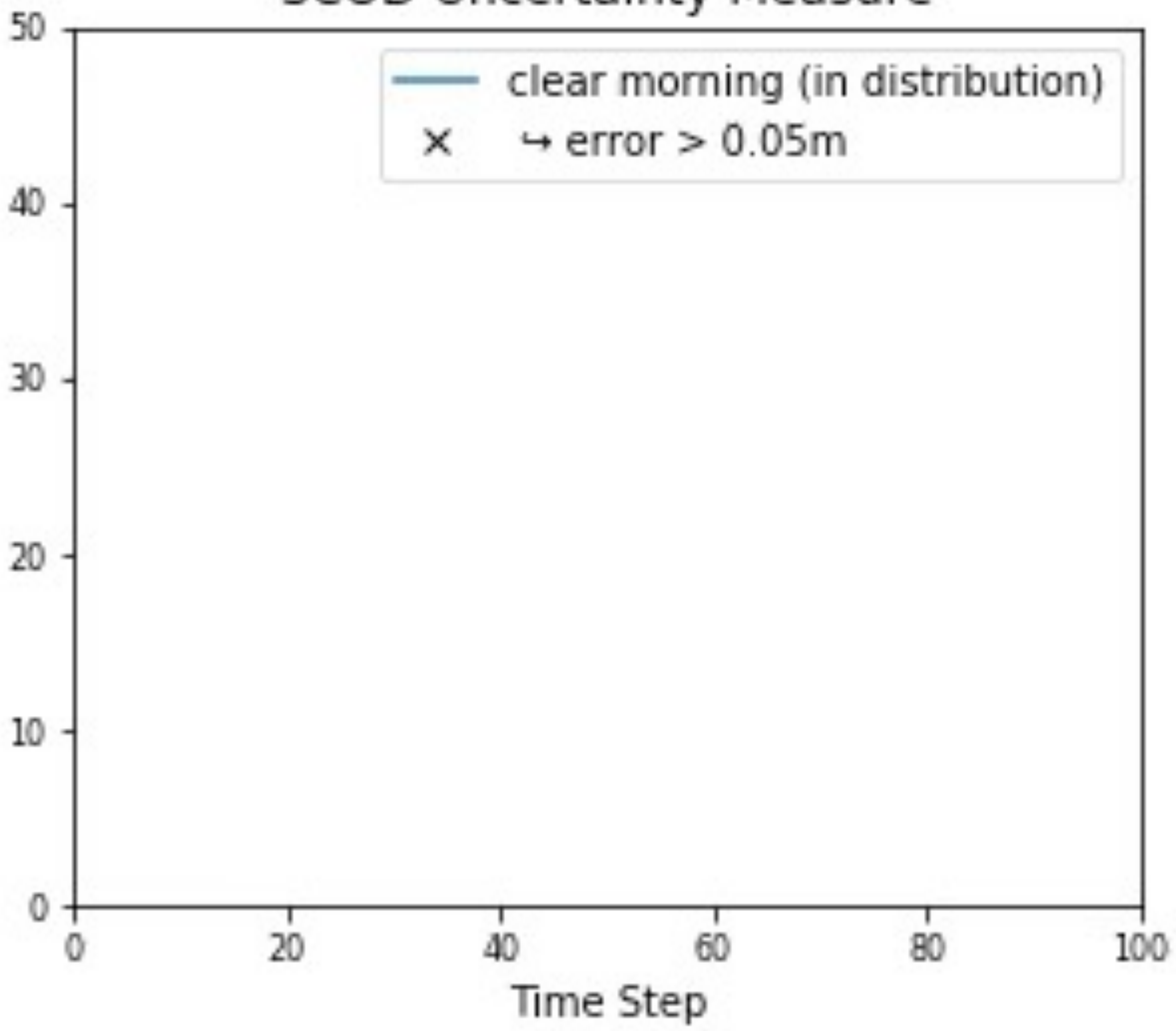
Tested on varying weather conditions and times of day

Key questions:
- How SCOD's uncertainty estimate behave on out-of-distribution settings?
- How does the uncertainty estimate correlate with model error?

Copyright by the authors. Published by The Aerospace Corporation with permission.

Qualitative Case Study

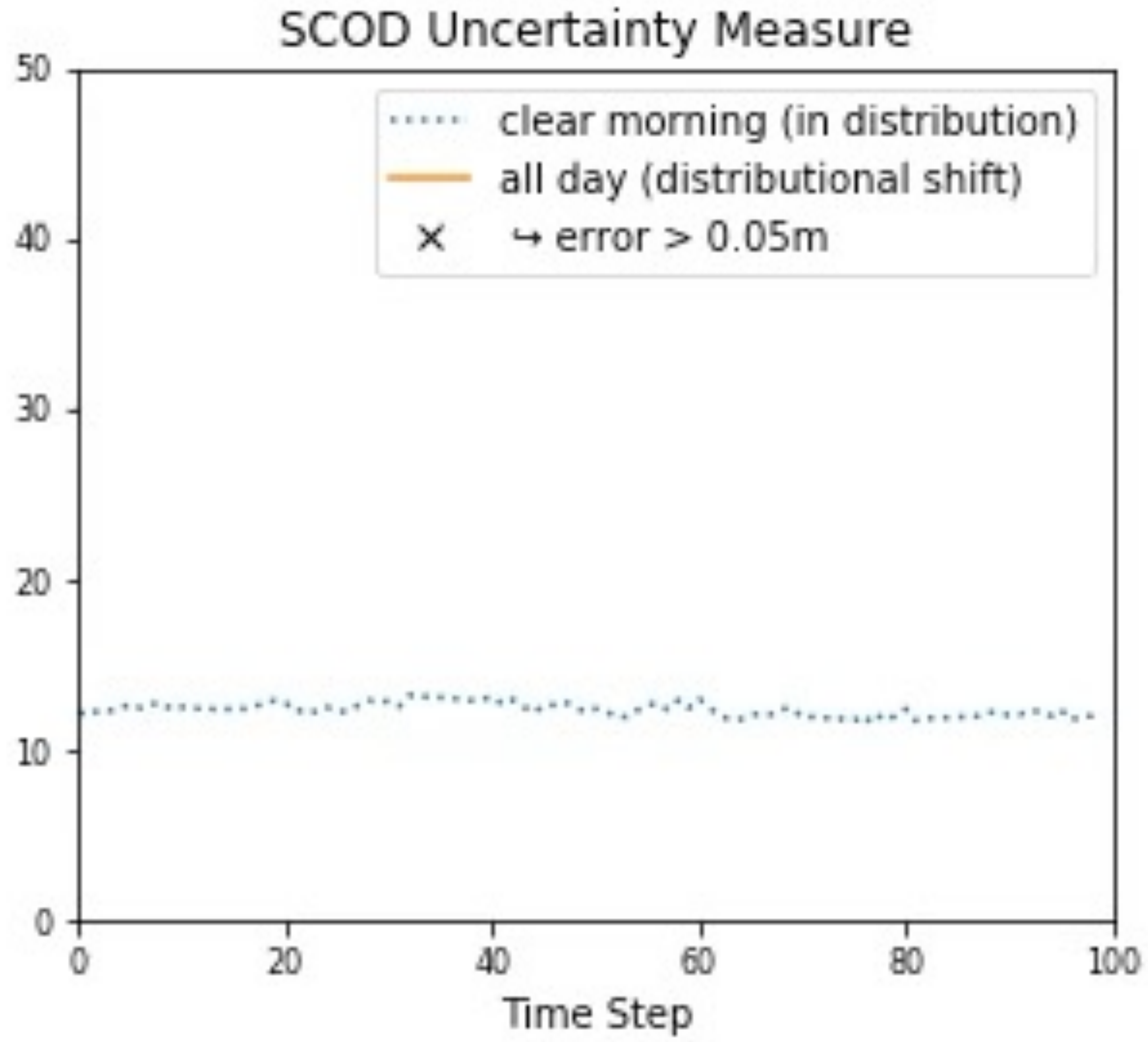
Visuomotor control of autonomous aircraft taxiing



Copyright by the authors. Published by The Aerospace Corporation with permission.

Qualitative Case Study

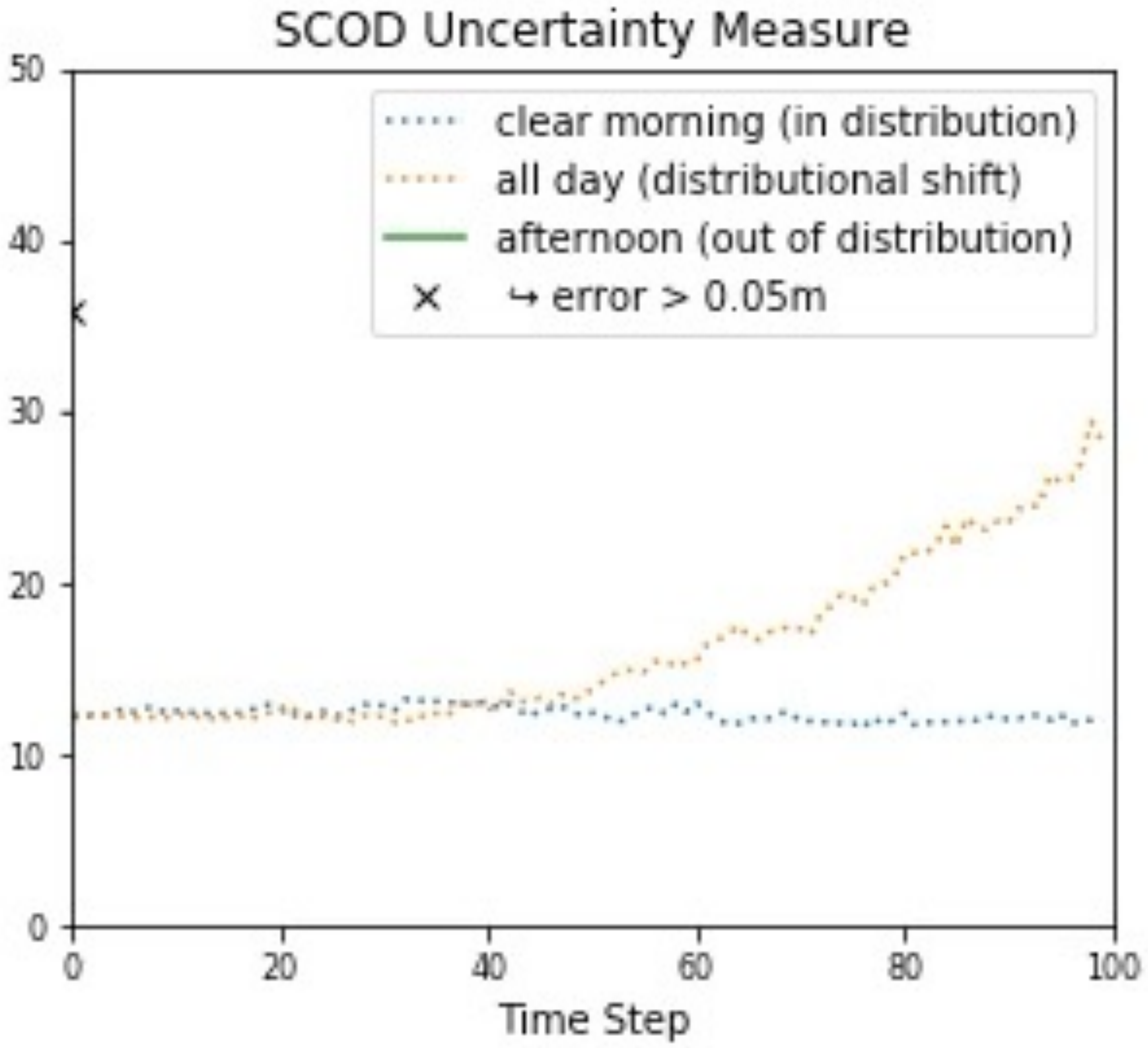
Visuomotor control of autonomous aircraft taxiing



Copyright by the authors. Published by The Aerospace Corporation with permission.

Qualitative Case Study

Visuomotor control of autonomous aircraft taxiing



Copyright by the authors. Published by The Aerospace Corporation with permission.

Quantitative Results

Performance in classifying OoD inputs (AUROC)

Compared against:

- ***Naïve***: use base DNN for uncertainty estimate
- General post-training uncertainty quantification methods:
  - ***Local Ensemble*** [Madras et al., 2019]
    Low-rank Hessian approx. computed via 2nd-order autodifferentiation
  - ***KFAC Laplace*** [Ritter et al., 2018]
    Layer-wise Kronecker-factored Hessian approx., sampled posterior at test time
- ***Deep Ensemble*** [Lakshminarayanan et al., 2018] (retrain K=5 identical models)

Copyright by the authors. Published by The Aerospace Corporation with permission.

Quantitative Results

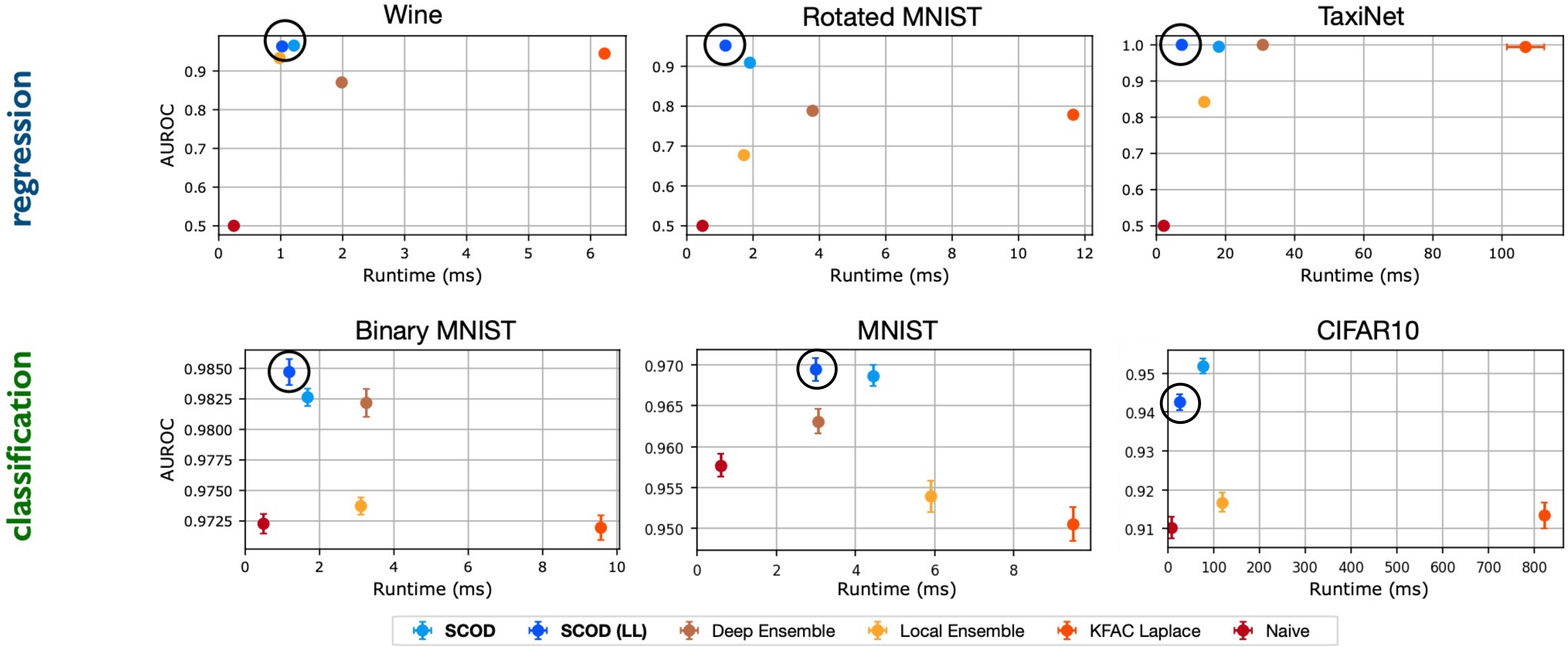
On a wide range of regression and classification tasks

	Experiment	In Dist	Out of Dist	Network
regression	**Wine** Properties -> Quality	Red wines	White wines	3 Layer MLP (11.4k params)
regression	**Rotated MNIST** Image -> Angle			3 Layer CNN (16.9k params)
regression	**TaxiNet** Image -> CTE and Heading			ResNet18 (11.2M params)
classification	**Binary MNIST** Image -> 0/1			3 Layer CNN (14.3k params)
classification	**MNIST** Image -> Digit			3 Layer CNN (15.5k params)
classification	**CIFAR 10** Image -> Class			DenseNet (7M params)

Copyright by the authors. Published by The Aerospace Corporation with permission.

Quantitative Results

Across a suite of regression and classification tasks, SCOD outperforms methods applicable to pre-trained models in terms of AUROC and runtime

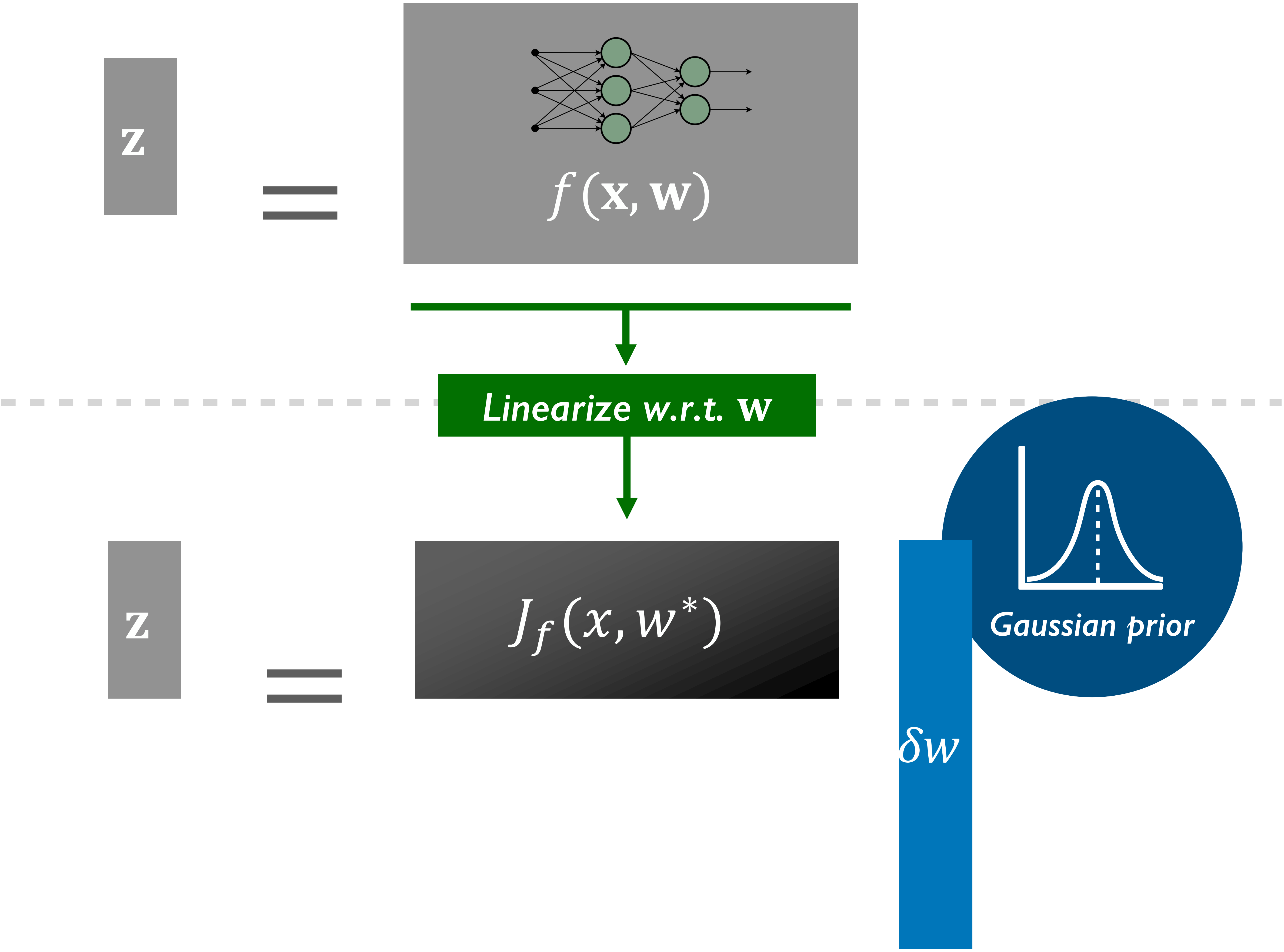


Copyright by the authors. Published by The Aerospace Corporation with permission.

Sketching Curvature for Efficient OOD Detection for Deep Neural Networks

SCOD was presented at UAI 2021, available on arXiv:2102.12567

Code is available at https://github.com/StanfordASL/SCOD/



Copyright by the authors. Published by The Aerospace Corporation with permission.

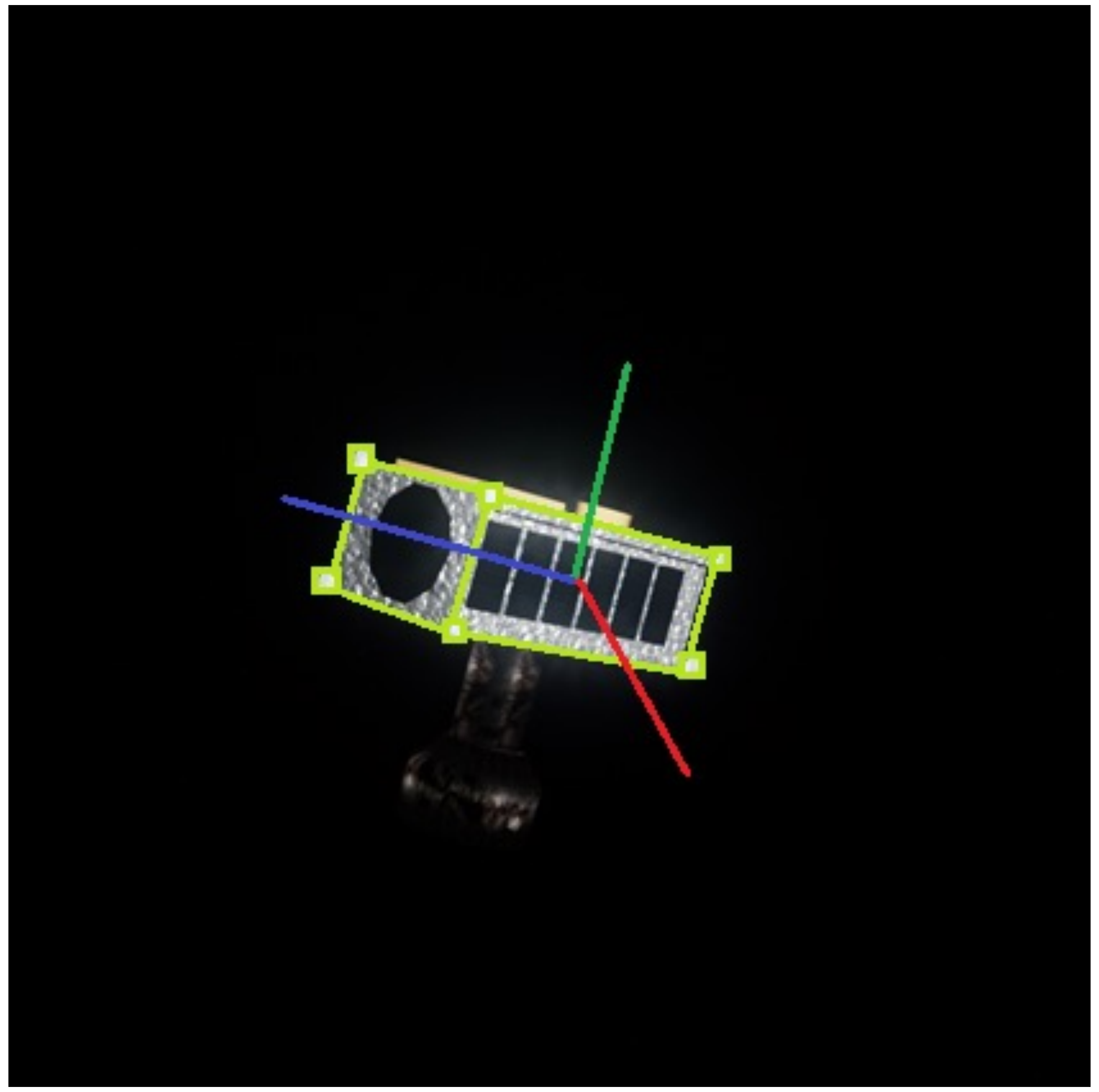
Outline

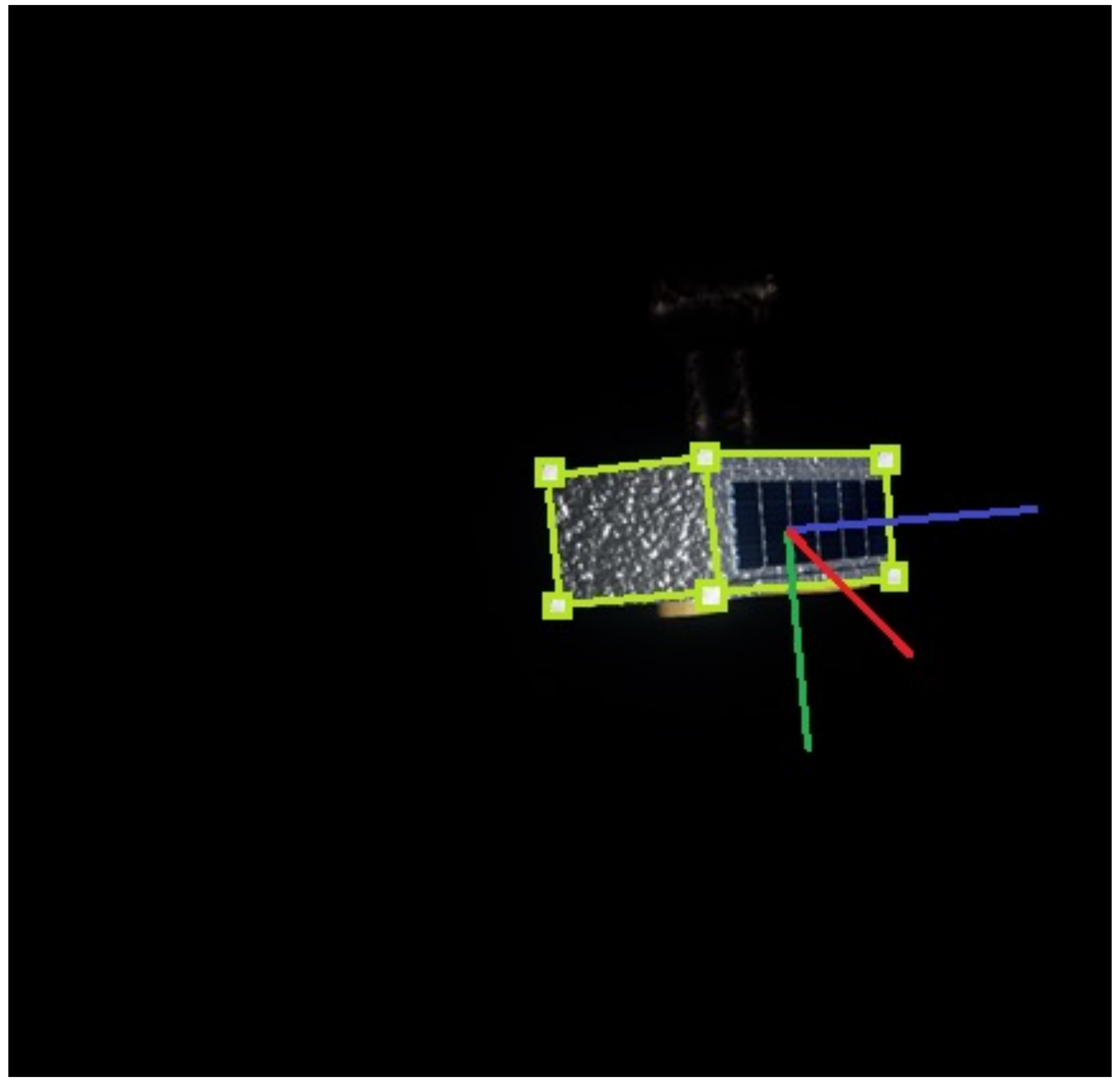
- Introduction
- Efficient anomaly detection for pre-trained DNNs
  - Problem Setup
  - SCOD: Sketching Curvature for Out-of-Distribution Detection
  - Insights and Results
- **Leveraging out-of-distribution detection in the data-collection pipeline**

Copyright by the authors. Published by The Aerospace Corporation with permission.

Future work: Efficient OOD detection for data labeling

Case study: ExoRomper dataset



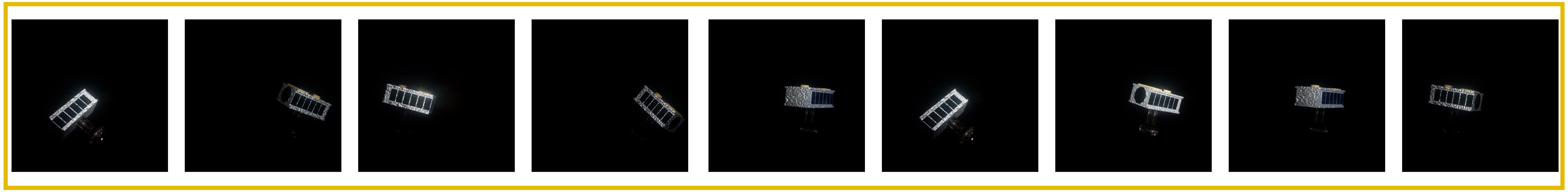


From an image, use a trained model to estimate pose (location + attitude) of a spacecraft

Copyright by the authors. Published by The Aerospace Corporation with permission.

OOD detection can identify areas where current DNN is not competent

Training dataset

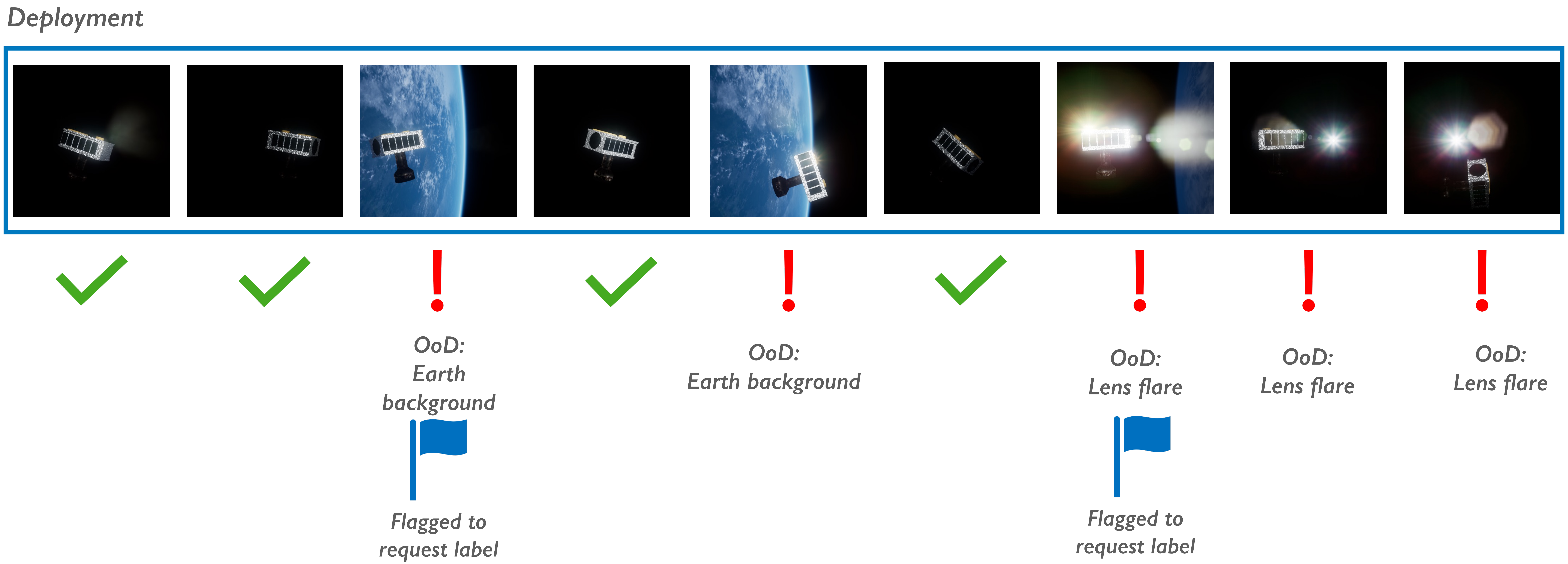


Deployment

OoD OoD OoD OoD OoD

Copyright by the authors. Published by The Aerospace Corporation with permission.

Can we use OOD information to select inputs to store and label for retraining?



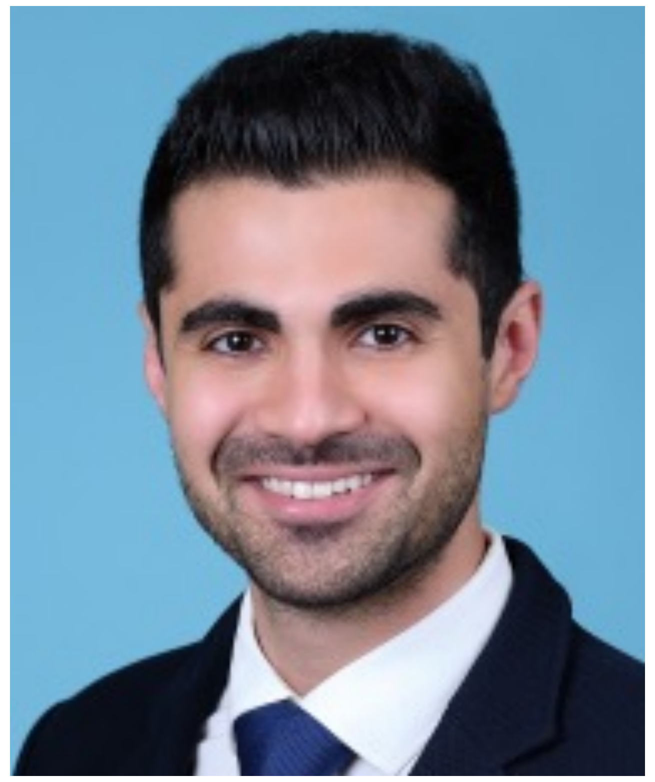
Goal: improve DNN performance while being cognizant of the costs of data storage and labeling

Copyright by the authors. Published by The Aerospace Corporation with permission.

Questions?

Somrita Banerjee

Ed Schmerling



Navid Azizan

Marco Pavone

Copyright by the authors. Published by The Aerospace Corporation with permission.

Copyright by the authors. Published by The Aerospace Corporation with permission.