Intersection Safety Systems (ISS)

Insights from the U.S. DOT Intersection Safety Challenge Stage 1B

U.S. Department of Transportation
ITS Joint Program Office

March 26, 2026

Source: iStock.com/Vertigo3d/cherzoff

Norah Ocel, P.E., PMP

Program Manager, Strategic Technology for Roadway Safety

Intelligent Transportation Systems Joint Program Office (ITS JPO)

U.S. Department of Transportation
ITS Joint Program Office

Source: Shutterstock.com/Marian Salabai

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers names appear in this presentation only because they are considered essential to the objective of the presentation. They are included for informational purposes only and are not intended to reflect a preference, approval, or endorsement of any one product or entity.

Except for the statutes and regulations cited, the contents of this presentation do not have the force and effect of law and are not meant to bind the States or the public in any way. This presentation is intended only to provide information regarding existing requirements under the law or agency policies.

This presentation was created and is being co-presented by both FHWA and Noblis. The views and opinions expressed in this presentation are the presenters' and do not necessarily reflect those of FHWA or the U.S. Department of Transportation (USDOT). The contents do not necessarily reflect the official policy of the USDOT.

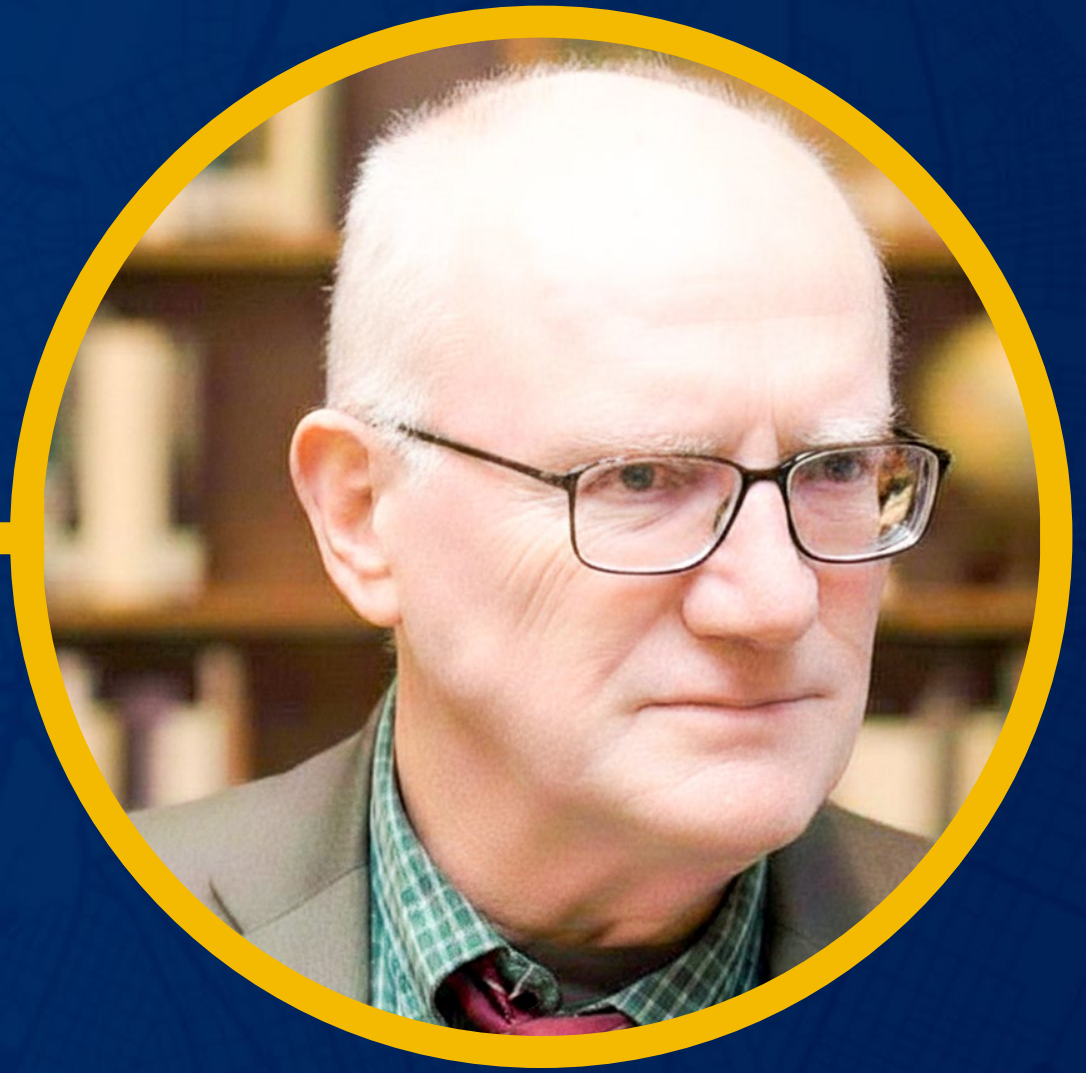
Purpose and Agenda

PURPOSE: Summarize results and insights from the U.S. DOT Intersection Safety Challenge Stage 1B: System Assessment and Virtual Testing in preparation for future ISS prototyping.

AGENDA:

- U.S. DOT Intersection Safety Challenge Overview
- Stage 1B Technical Evaluation: Methodology
- Stage 1B Technical Evaluation: Results
- Key Insights from the Challenge
- Future Direction
- Q&A

Chris Atkinson, Sc.D.
Fellow SAE, Fellow ASME



Deputy Director for Technology

Advanced Research Projects Agency – Infrastructure (ARPA-I)

Office of the Assistant Secretary for Research and Technology (OST-R)

U.S. Department of Transportation
ITS Joint Program Office

Overview

U.S. DOT Intersection Safety Challenge

Image Source: U.S. DOT

U.S. DOT Intersection Safety Challenge Overview

VISION: Transform intersection safety through the innovative application of emerging technologies including machine vision, sensor fusion, and real-time decision-making to identify and mitigate unsafe conditions involving vehicles, pedestrians, and other road users.

PRIZE COMPETITION: Encourage teams of innovators to develop and virtually test their intersection safety systems to compete for prizes.

STRUCTURE:

Prize Competition (Stage 1A and Stage 1B)

✓ **Stage 1A:**
Concept Assessment
15 winners
(FY23-24)

✓ **Stage 1B:**
System Assessment & Virtual Testing
10 winners
(FY24-25)

Intersection Safety Systems (ISS) Concept

SENSE

Sense the intersection conditions using emerging, low-cost sensors (e.g., visual camera, LiDAR, radar, thermal camera) and data fusion

THINK

Use artificial intelligence (AI) to improve situational awareness and predict safety threats

ACT

Mitigate unsafe conditions involving vehicles and other road users by issuing warnings and/or modifying signalized control settings

INTERSECTION SAFETY CHALLENGE

ISS can help identify and mitigate conflicts.

Image Source: U.S. DOT

INTERSECTION SAFETY CHALLENGE

In **Stage 1B: System Assessment and Virtual Testing**, the teams' algorithms were assessed against three key technical elements of ISS operations via a data science competition.

Image Source: U.S. DOT

SENSE

THINK

ACT

TECHNICAL ELEMENTS OF ISS ASSESSED IN STAGE 1B:

1) Detection, Classification, Localization
2) Path Prediction
3) Conflict Prediction

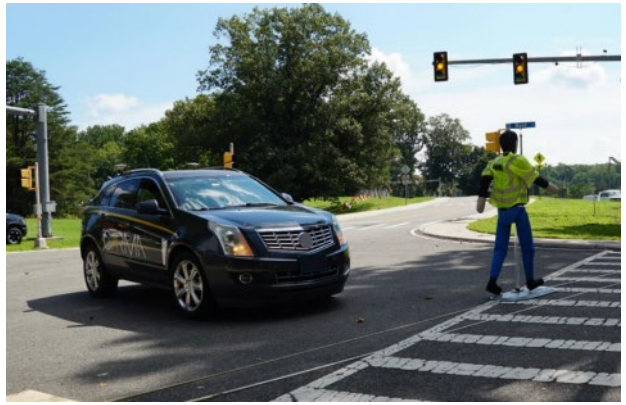
U.S. Department of Transportation
ITS Joint Program Office

Stage 1B Data Collected at FHWA Test Facility

The multi-sensor competition data include potential conflict-based and non-conflict-based experimental scenarios between pedestrians, bicyclists, and vehicles during both daytime and nighttime conditions.*

Sensor Type	Number of Sensors
Closed-circuit television (CCTV) visual camera	8
Thermal camera	5
Radar sensor	4
Light detection and ranging (LiDAR) sensor	2

**Note that no human road users were put at risk of being involved in a collision during the data collection efforts.*



Conflict scenario between a surrogate adult pedestrian and vehicle

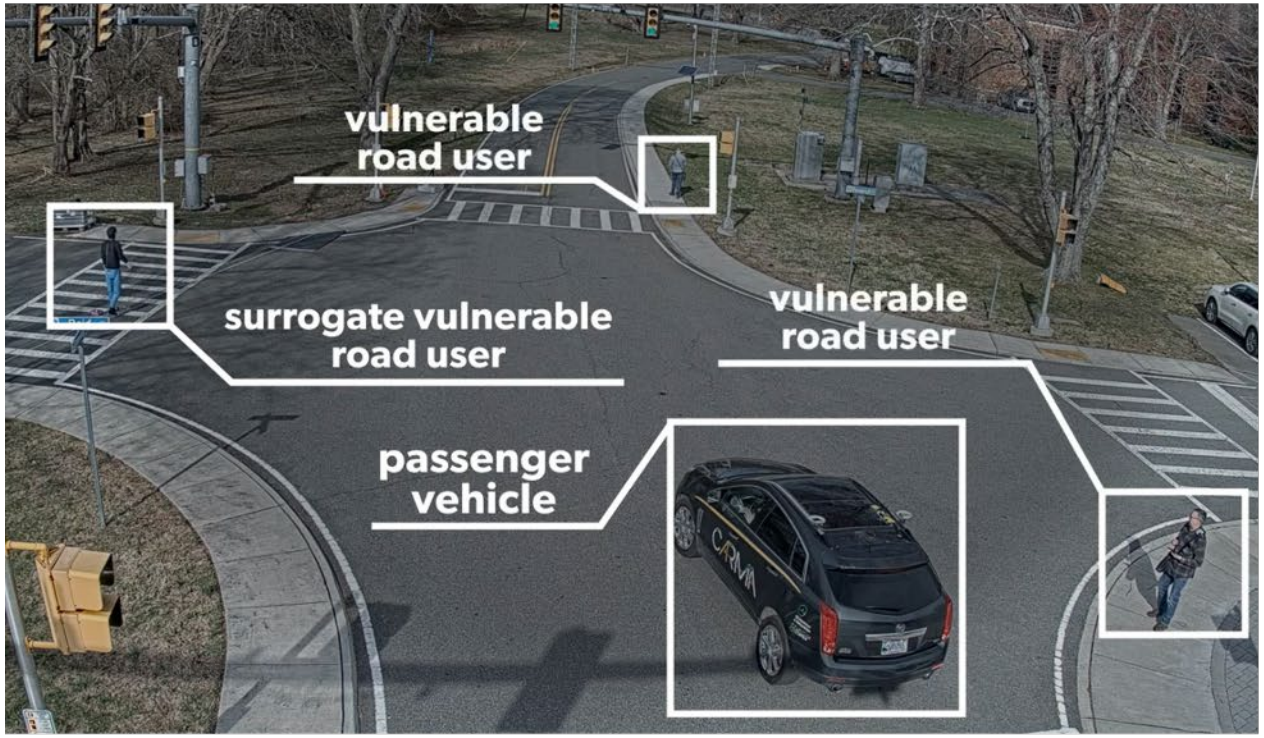
Surrogate pedestrian and bicyclist testing devices for high-risk crash scenarios

Example props potentially used by pedestrians in the data

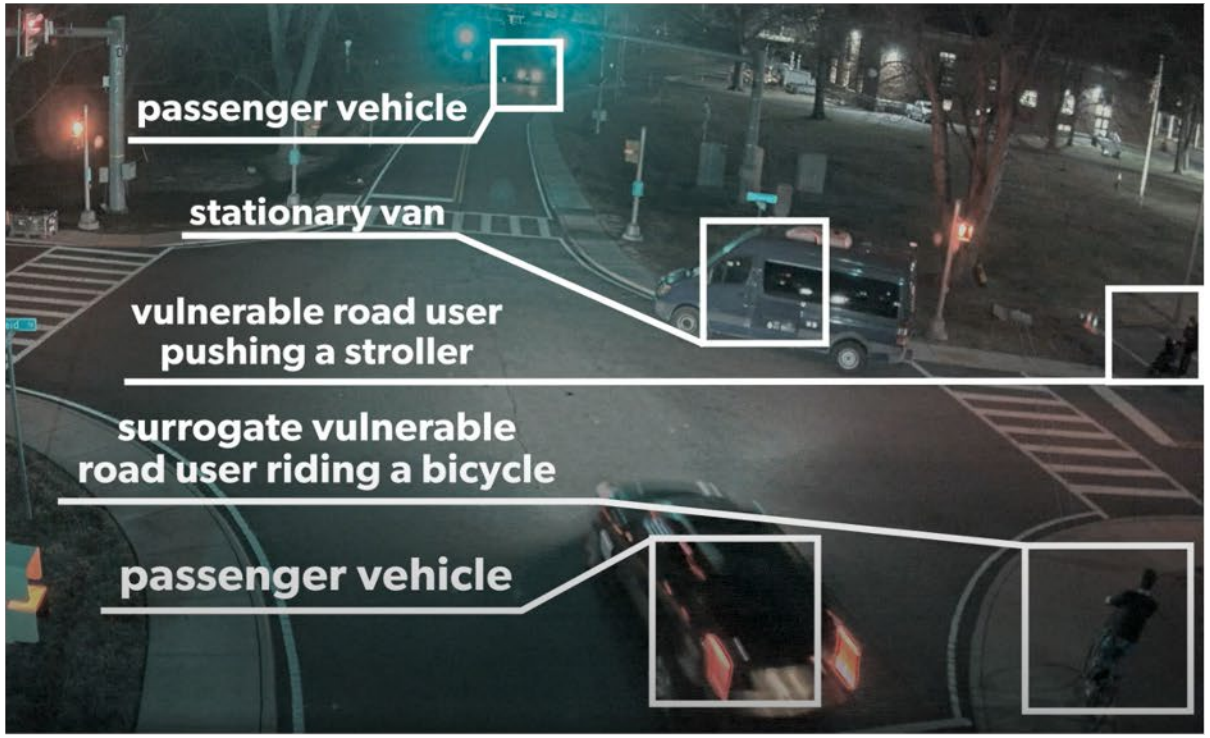
Images Source: FHWA

Stage 1B Data: Example Scenarios

Daytime Example



Nighttime Example



Images Source: FHWA

Jesse Eisert, Ph.D.

Research Psychologist

Office of Research, Development, and Technology

Federal Highway Administration (FHWA)

U.S. Department of Transportation
ITS Joint Program Office

Stage 1B Evaluation: Methodology

U.S. DOT Intersection Safety Challenge

Image Source: U.S. DOT

Deriving Ground Truth Data

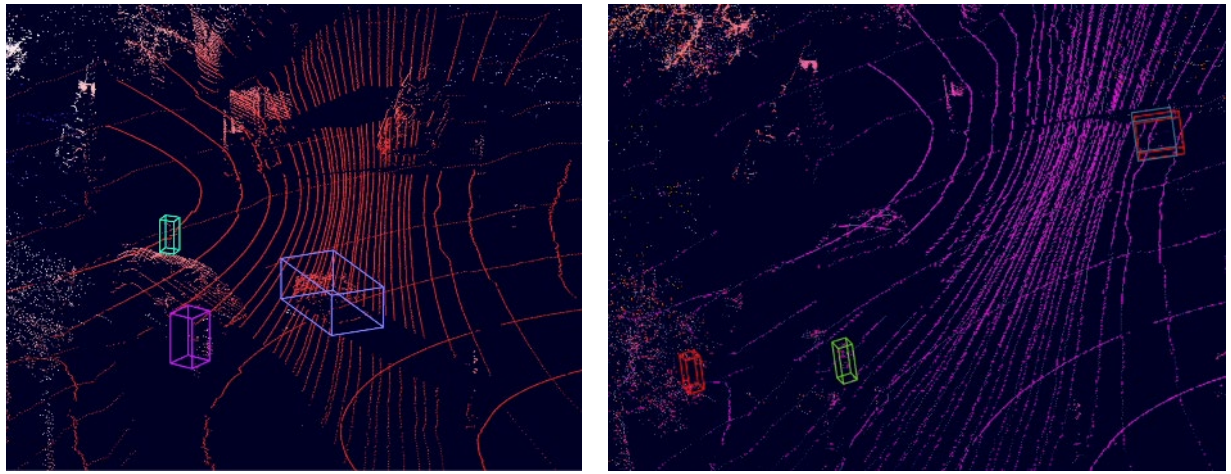
3 Datasets:

- Participants received training (unlabeled) and validation (labeled) sensor data to prepare their algorithms and understand submission formats.
- Test dataset
  - Participants received unlabeled raw sensor data cutoff before the potential conflict to prevent manual labeling.
  - U.S. DOT reserved the ground truth labels for evaluation.

Dataset	Raw Sensor Data	Ground Truth Labels
Training	☑	N/A
Validation	☑	☑
Test	Cutoff before potential conflict	Withheld from Participants

Ground Truth Data:

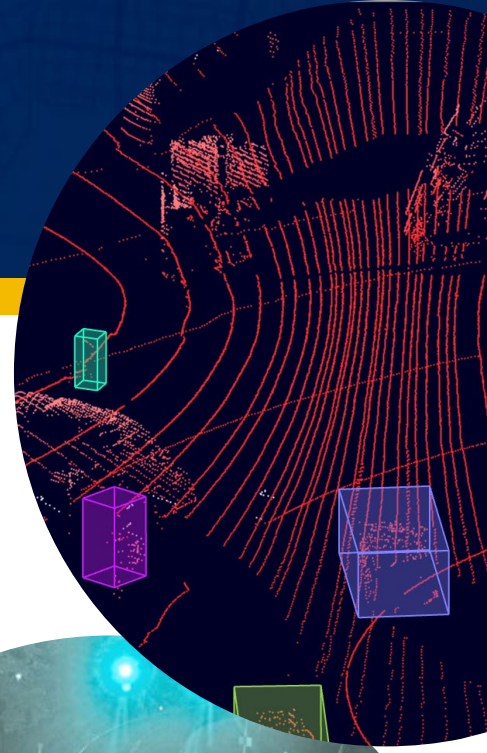
- Position Data
  - 3D Coordinate Data (x, y, z)
  - 3D Bounding Box Data (length, width, height)
  - Yaw (ranging from 0° to 359°)
- Road User Label
- Conflict/No Conflict Label

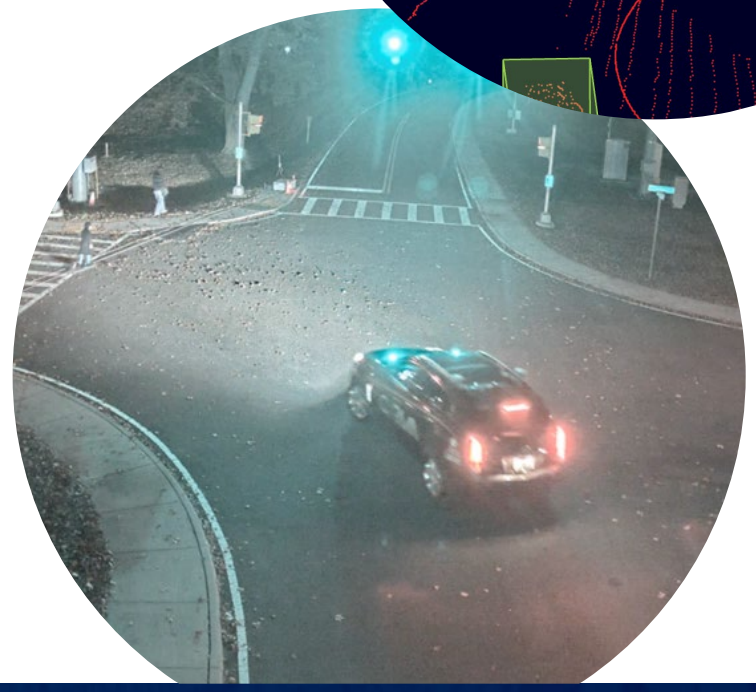


Images Source: U.S. DOT

Example 3D Ground Truth Bounding Boxes

Evaluation Metrics for ISS Technical Elements





Detection, Classification, and Localization

- Mean Average Precision (mAP) – Enables holistic bounding box evaluation across several precision thresholds
- Captured both coarse (class) and fine (subclass) classification information for comprehensive evaluation

Path Prediction

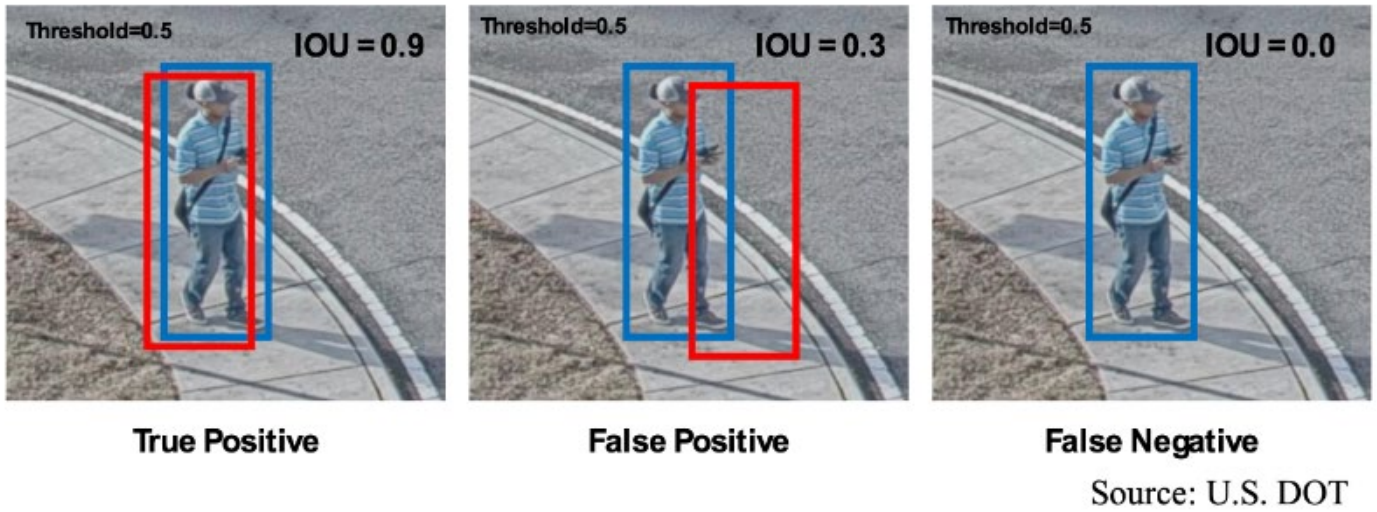
- mAP
- Average Displacement Error (ADE) – the average Euclidean distance between the predicted and actual location of a road user

Conflict Prediction

- F2 Score – Captures both precision and recall, but prioritizes proportion of actual positives (conflicts) that were correctly identified
- False negatives (missed conflicts) are considered a greater safety risk than false positives.

Images Source: U.S. DOT

Mean Average-Precision (mAP)



Detection, Classification, and Localization (DCL) were collectively evaluated using the **mAP_2** at different **intersection over union $(IoU)_1$** thresholds.

Formulas and Calculations

$$IoU(G,P) = \frac{G \cap P}{G \cup P}$$

[1]**Intersection over Union (*IoU*) calculation**: ratio of intersecting predictive and true bounding boxes over non-intersection

$$mAP = \frac{1}{C}\frac{1}{10}\Sigma_c\Sigma_{a\in\alpha}\, AP_{c,a}$$

[2]**Mean Average Precision**: Average Precision over all classes across ten thresholds (α)

Average Displacement Error (ADE)

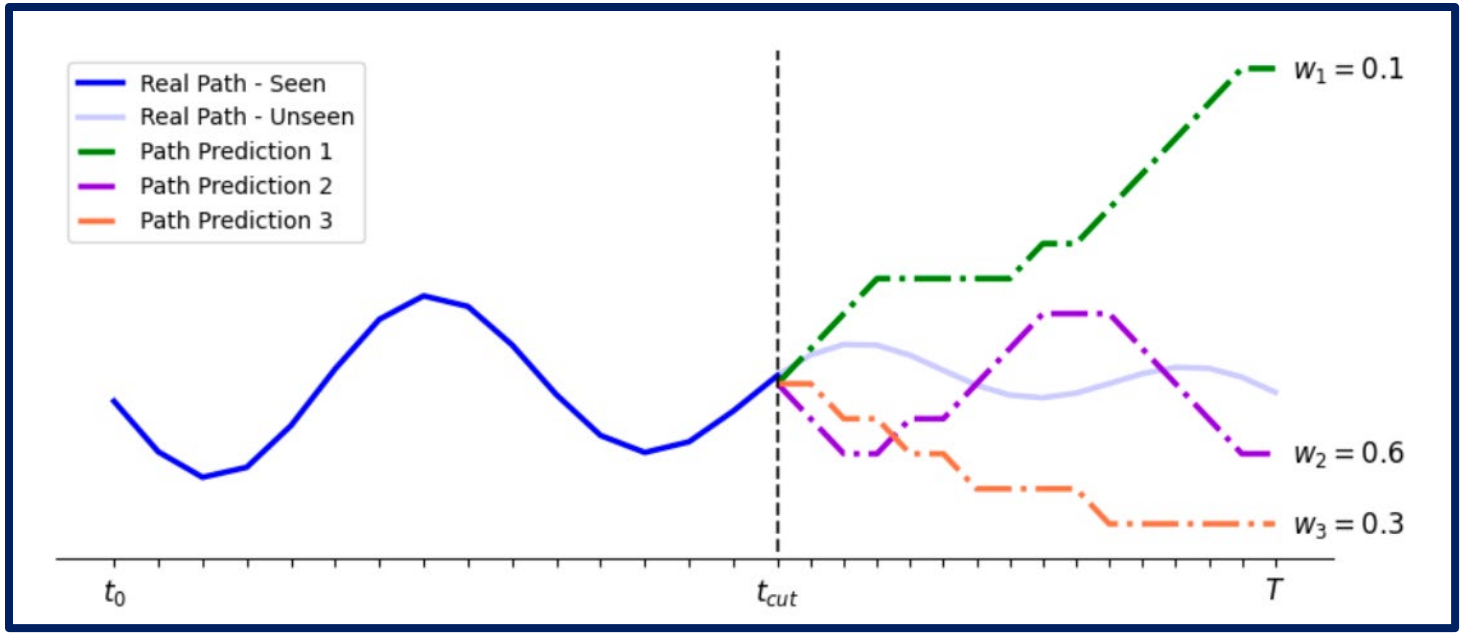


Image Source: U.S DOT

Path Prediction performance was evaluated using the **ADE_1** metric.

Positional accuracy is calculated as an average across up to three candidate paths weighted by probability.

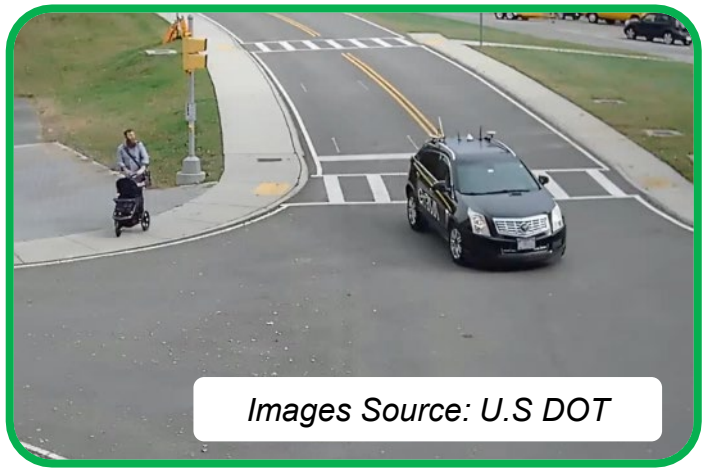
Formulas and Calculations

$$\Sigma_c I_c(c) * \frac{1}{T - t_{cut}} \Sigma_{t=t_{cut}+1}^{T} \sqrt{(x_{c,t} - \Sigma_i(w_{c,i} * \hat{x}_{c,i,t})^2 + (y_{c,t} - \Sigma_i(w_{c,i} * \hat{y}_{c,i,t}))^2 + (z_{c,t} - \Sigma_i(w_{c,i} * \hat{z}_{c,i,t}))^2}$$

[1]**ADE**: the Euclidean or L2 distance between the prediction and the ground truth, averaged over all timestamps

F2 Conflict Score

Conflict between road users



Non-Conflict between road users

Conflict prediction between road users was evaluated using the **$F2_2$ score**.

Formulas and Calculations

$$F_2 = \frac{5TP}{5TP + 4FN + FP}$$

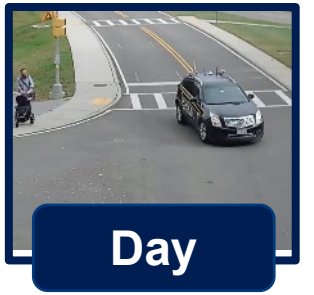
[1]**Modified F2 Score**: F2 score typically balances recall (*penalizes false negatives*) and precision (*evaluates false positives*)

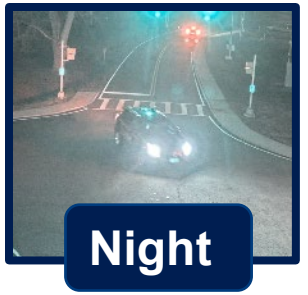
Due to the severity of missing roadway conflicts, **we modify F_2 to emphasize false negatives**

Operational Conditions

DCL, Path Prediction, and Conflict-F2 performance were also analyzed by **operational condition**: *particular scene settings* (day vs. night) or *road user behavior* (vehicle left-turn vs. right-turn). Evaluation across these settings enables the identification of weak spots for state-of-the-art ISS.

- **Conflict / Non-Conflict:** Runs containing a conflict between two road users vs. runs without conflicts
- **Occlusion / Non-Occlusion:** Runs in which one or more sensors' views are obstructed by a large object or vehicle vs. runs in which the entire intersection is visible
- **Left-Turn / Right-Turn:** Runs in which a vehicle makes a left-turn or right-turn in the observed intersection
- **Day / Night:** Runs taking place during daylight hours vs. runs taking place at night





Claire Silverstein, PhD, PMP

Senior Principal, Surface Transportation Systems

Federal Civilian Solutions

Noblis

U.S. Department of Transportation
ITS Joint Program Office

Stage 1B Evaluation: Results

U.S. DOT Intersection Safety Challenge

Image Source: U.S. DOT

Classification: Overall mAP Distribution

Image Source: U.S. DOT

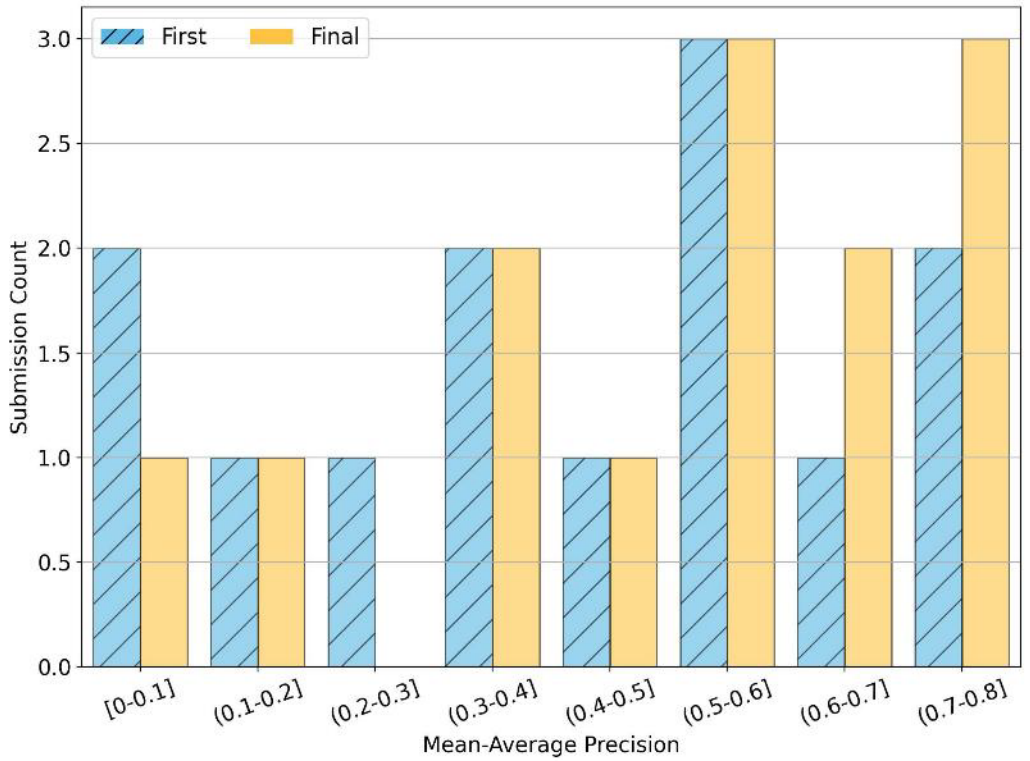


Figure: Histogram of classification mean Average Precision results

Note on First vs. Final: First evaluations capture performance from teams with minimal time to analyze the dataset (72 hours), while **Final** evaluations capture outcomes after familiarization (6 weeks).

Result Statistics

- The range defines a 0.65 mAP spread (0.14 minimum to 0.79 maximum).
- The mean score and variance are 0.55 and 0.04, respectively.

Path Prediction: Overall ADE Distribution-1

Image Source: U.S. DOT

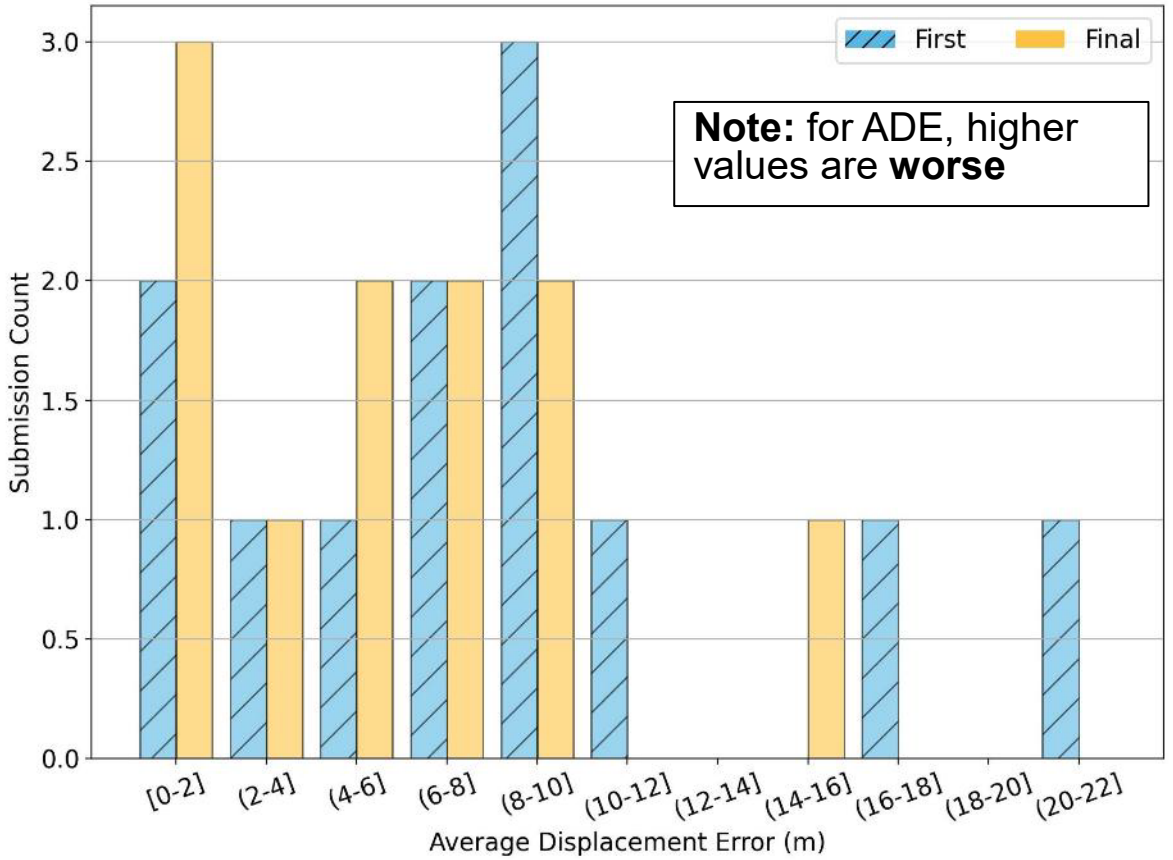


Figure: Histogram of Average Displacement Error scores for initial and final evaluations.

Result Statistics

- Final ADE scores fell primarily in the 2.21 m–9.01 m range (6.8 m spread), except for one 15.7 m ADE outlier (13.49 m spread).
- The top 4 submissions' ADEs are bounded by 4 m with a range of 2.21 m–3.81 m.

Path Prediction: Overall ADE Distribution-2

Image Source: U.S DOT

The path prediction performance of the top four ISS is roughly 4 meters of error. **This is nearly the length of one vehicle.**

Result Statistics

- Final ADE scores fell primarily in the 2.21 m–9.01 m range (6.8 m spread), except for one 15.7 m ADE outlier (13.49 m spread).
- The top 4 submissions' ADEs are bounded by 4 m with a range of 2.21 m–3.81 m.

Conflict Prediction: Overall Conflict-F2 Distribution

Image Source: U.S. DOT

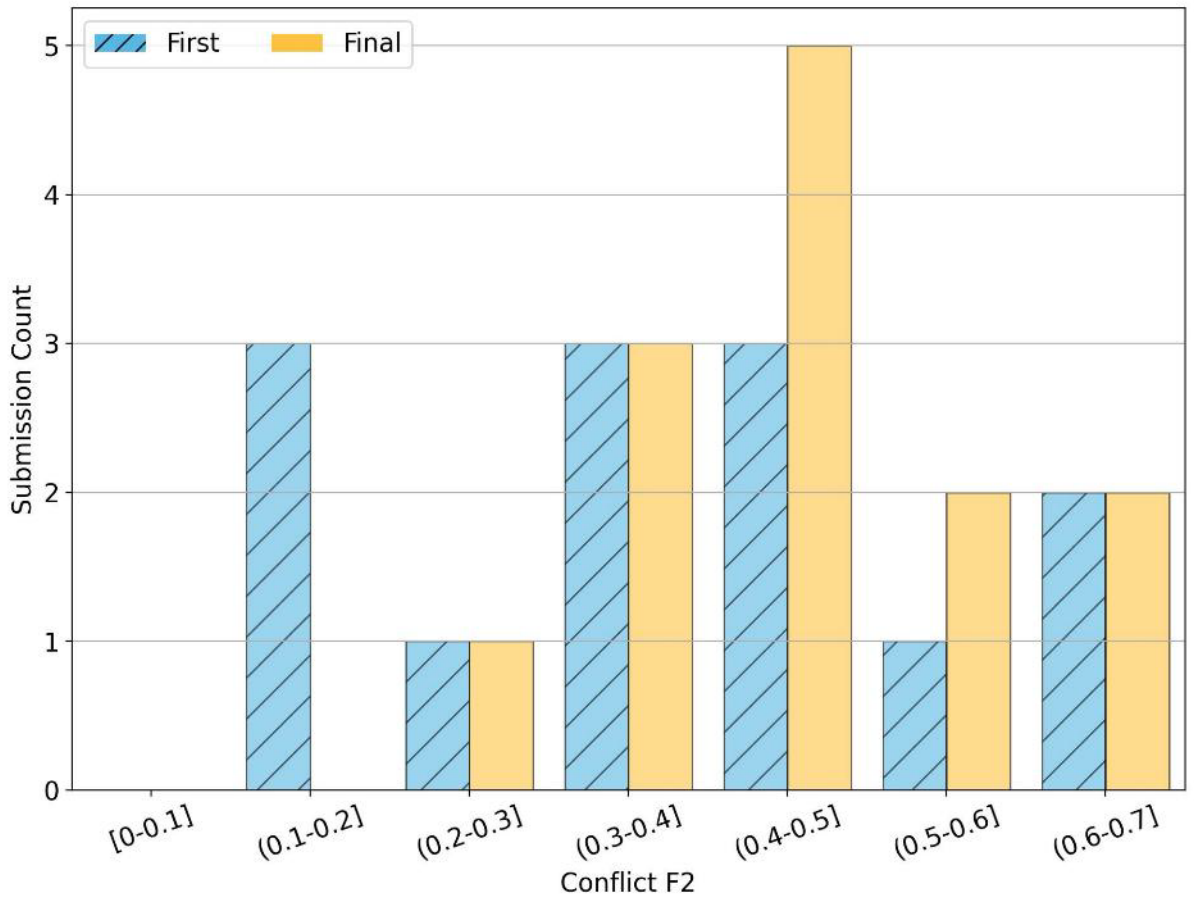


Figure: Histogram of Conflict F2 scores for initial and final evaluations.

Result Statistics

- Submitted results define a 0.25–0.67 range (0.42 spread).
- The mean score and variance are 0.47 and 0.03, respectively.

Discussion

F2 Scores are much more concentrated around the mean than mAP or ADE metrics.

Classification: Overall mAP by Subclass

Image Source: U.S. DOT

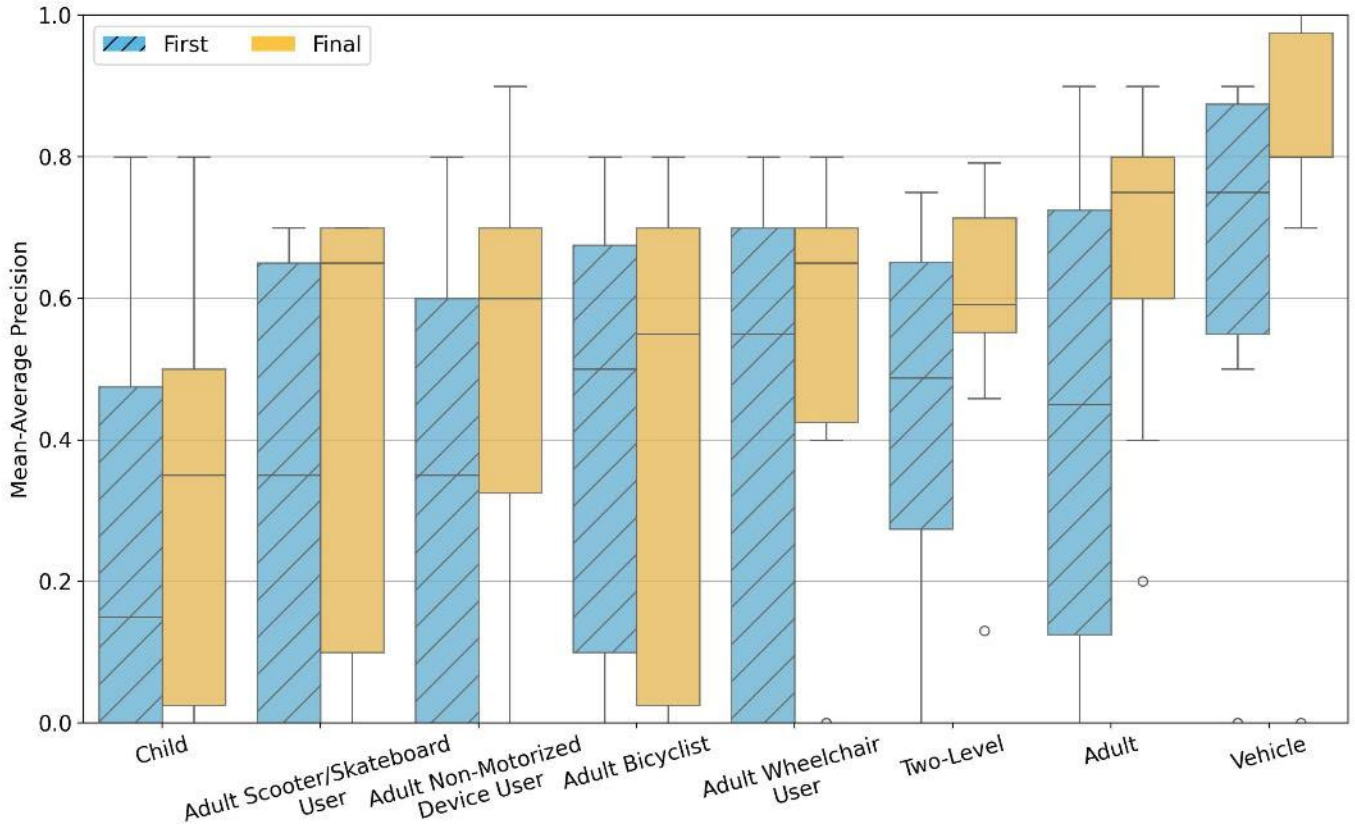


Figure: Box plot of mAP scores for first and final evaluations by Subclass.

Result Statistics

- On average, **mAP scores improved by 10% (~0.1 mAP)** between first and final submissions
- ISS detect, classify, and locate, **Vehicles most accurately** and **Child subclasses least accurately**
- Performance is positively correlated with representation in the dataset

Path Prediction: Overall ADE by Subclass

Image Source: U.S. DOT

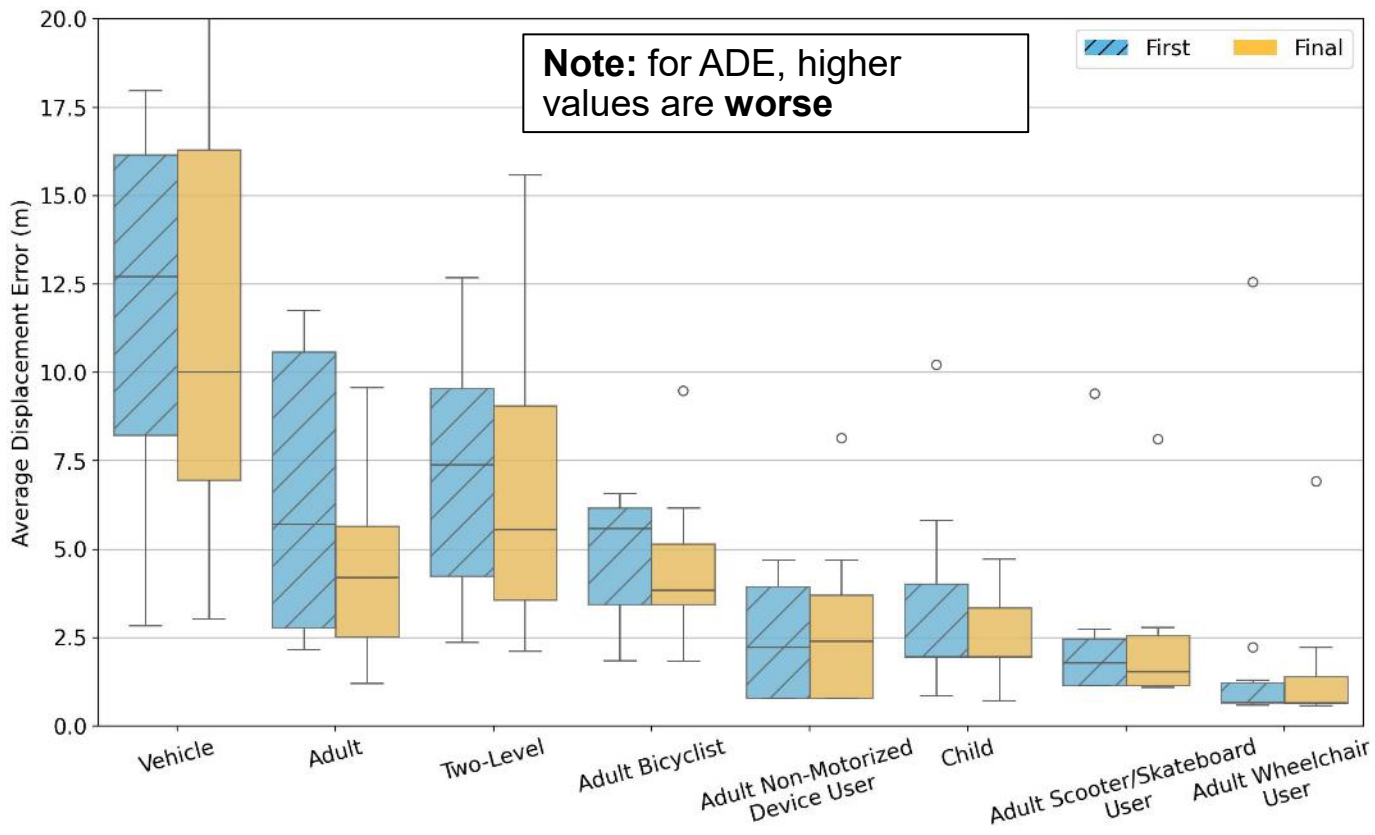


Figure: Box plot of mAP scores for first and final evaluations by Subclass.

Result Statistics

Subclasses perform in nearly inverse order to that of their mAP performance. Vehicle performs **~6 meters worse** than the next worst road user.

Discussion

This may be partially misleading, since vehicles move further distances than pedestrian or cyclist road users.

Classification: Operational Conditions by mAP

Final Detection, Classification, and Localization Scores by Run Scenario

Scenario Groups	Count of Runs by Scenario Group	mAP Score (Avg.)	mAP Score (Avg.) Subsampled (20 runs)
Non-Conflict	144	0.55	0.39
Day	100	0.54	0.38
Right-Turn	81	0.53	0.38
Non-Occlusion	85	0.52	0.37
Left-Turn	86	0.51	0.36
Night	85	0.50	0.35
Occlusion	60	0.48	0.34
Conflict	41	0.47	0.33

Discussion

Teams performed modestly better on **Right-Turn**, **Day**, and **Non-Occlusion**, and **Non-Conflict** runs in comparison to their respective counterparts (Left-Turn, Night, Occlusion, and Conflict).

Path Prediction: Operational Conditions by ADE

Final Path Prediction Error by Run Scenario

Scenario Groups	Count of Runs by Scenario Group	ADE Score (Avg.) (m)	ADE Score (Avg.) (m) Subsampled (20 runs)
Right-Turn	81	5.42	8.20
Day	100	6.44	9.43
Non-Conflict	144	6.48	7.78
Occlusion	60	6.51	8.82
Non-Occlusion	85	6.54	9.22
Left-Turn	86	6.65	9.31
Night	85	6.67	8.80
Conflict	41	6.81	7.88

Note: for ADE, higher values are **worse**

Result Statistics

Generally, **the range of scores by scenario is tighter than by** subclass, but relative rank is preserved after subsampling).

Discussion

Teams performed modestly better on **Right-Turn, Day,** and **Non-Conflict** runs in comparison to their respective counterparts (**Left-Turn, Night, and Conflict**).

Conflict Prediction: Operational Conditions by F2

Final Conflict F2 by Run Scenario

Scenario Groups	Number of Runs by Scenario Group	Conflict F2 (Avg.)
Right-Turn	81	0.56
Occlusion	60	0.56
Day	100	0.50
Night	85	0.49
Left-Turn	86	0.45
Non-Occlusion	85	0.37

Result Statistics

Most run group scores fall within a relatively tight range (0.56–0.45); however, unlike Classification & Path Prediction, one group performs as a low outlier (Non-Occlusion).

Discussion

Teams performed better on Right-Turn and Occlusion runs.

Results by Reported Sensors Used (#1 of 2)

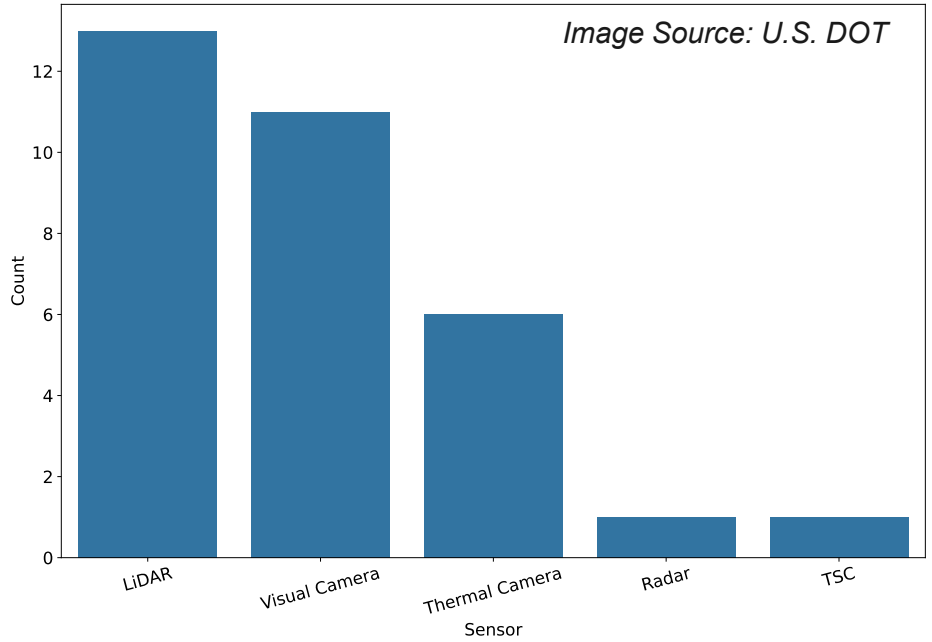


Figure: Histogram of reported sensors used.

LiDAR and Visual Cameras were the **most frequently used sensors** and were the most common sensor pair.

Final Detection, Classification, and Localization Scores by Reported Sensor Used

Sensor	# Teams Used	Overall Avg. mAP Score	Day / Night Avg. Difference	Non-Occlusion/Occlusion Avg. Difference
LiDAR	11	0.406	+0.030	+0.024
Visual Camera	11	0.497	+0.036	+0.030
Thermal Camera	6	0.445	+0.020	+0.028
Radar	1	0.672	+0.081	+0.086

Final Path Prediction Error by Reported Sensor Used

Sensor	# Teams Used	Overall Avg. ADE Score (m)	Day / Night Avg. Difference (m)	Occlusion/Non-Occlusion Avg. Difference (m)
LiDAR	11	9.12	+0.632	+0.406
Visual Camera	11	7.77	+0.817	+0.395
Thermal Camera	6	7.50	+0.029	+0.110
Radar	1	2.38	-0.226	-0.086

Results by Reported Sensors Used (#2 of 2)

Images Source: U.S DOT

Teams reporting Visual Camera use saw (on average) the highest mAP scores, but also the highest day-to-night performance drops.

Final Detection, Classification, and Localization Scores by Reported Sensor Used

Sensor	# Teams Used	Overall Avg. mAP Score	Day / Night Avg. Difference	Non-Occlusion/Occlusion Avg. Difference
LiDAR	11	0.406	+0.030	+0.024
Visual Camera	11	0.497	+0.036	+0.030
Thermal Camera	6	0.445	+0.020	+0.028
Radar	1	0.672	+0.081	+0.086

Final Path Prediction Error by Reported Sensor Used

Sensor	# Teams Used	Overall Avg. ADE Score (m)	Day / Night Avg. Difference (m)	Occlusion/Non-Occlusion Avg. Difference (m)
LiDAR	11	9.12	+0.632	+0.406
Visual Camera	11	7.77	+0.817	+0.395
Thermal Camera	6	7.50	+0.029	+0.110
Radar	1	2.38	-0.226	-0.086

Key Insights

U.S. DOT Intersection Safety Challenge

Jesse Eisert, FHWA

Image Source: U.S. DOT

Challenge Structure Strengths and Limitations for assessing ISS

STRENGTHS

+ Focused on algorithms
+ Reduced calibration requirement from teams
+ Enabled consistent, quantitative scoring
+ Allowed granular performance comparisons
+ Protected participants' intellectual property (IP)

LIMITATIONS

- Did not allow for real-time (synchronous) system-in-the-loop assessment
- Limited the evaluation to a single intersection
- Restricted hardware innovation
- Restricted modeling flexibility (e.g., conflict definition)
- Limited in its ability to create and share new products

Overall Challenge Effectiveness

- **Task-based performance variability clarified technical maturity.** While foundational perception capabilities are advancing, more work is needed in prediction and decision support within ISS.
- **Challenge constraints successfully focused on algorithmic performance.** By providing the same multi-modal sensor inputs to all participants, the Challenge emphasized algorithmic performance over hardware tuning, making it possible to isolate core technical capabilities across different team solutions.
- **Performance gaps were uncovered.** The competition structure revealed important performance gaps for ISS, including child detection, left-turn scenarios, and nighttime conditions.
- **Results highlighted potentially optimal sensor fusion approaches for ISS.** Teams that employed multi-sensor fusion demonstrated higher performance compared to those that used a single sensor type.

The Stage 1B data science competition structure balanced fairness, technical rigor, and scenario complexity, allowing meaningful differentiation of each team's algorithmic capabilities.

Overall Key Takeaways from Stage 1B (#1 of 2)

- **Challenge results are encouraging for ISS prototyping.** Many teams were capable of suitably accurate detection, classification, and localization. Path and conflict prediction results were also promising, but the Stage 1B competition was limited in its ability to assess these ISS capabilities.
- **Teams struggled most in detecting and predicting movement for the surrogate child.** Lower performance on all metrics with respect to the Child road user subclass points to a high-risk need for further research, development, and testing.
- **Nighttime and vehicle left turns point to areas for ISS improvement.** Teams performed better on day and right-turn runs compared to night and left-turn runs respectively across all metrics.
- **Road user speed appeared to impact technical performance differently, depending on the metric.** Teams typically performed a little better on path prediction for fast road users compared to slower road users, perhaps since slow movement may indicate or precede a change in movement direction.

Image Source: U.S. DOT

U.S. Department of Transportation
ITS Joint Program Office

Overall Key Takeaways from Stage 1B (#2 of 2)

- **Sensor fusion outperformed individual sensor approaches.** Teams using two or more sensor types in their systems clearly performed better than teams relying on a single sensor type.
- **Teams recognized and adapted to data quality challenges.** Although teams identified inherent flaws and unexpected inconsistencies in the data (reflective of potential real-world situations), they adapted and used the imperfect data to build a robust ISS.
- **Teams acknowledged the value of the data challenge.** Teams viewed the challenge's premise as a crucial call-to-action and many outlined ambitions to extend their ISS work beyond this initiative.

Stage 1B Technical Lessons Learned

- **Anticipate and mitigate sensor calibration complexity early in the process.** Sensor calibration requires tailored methods for each modality and is labor-intensive. Reserve sufficient time for calibration and data quality checking.
- **Engage vendors early and secure data access agreements for proprietary sensors like radar.** In Stage 1B, working with radar data involved navigating proprietary restrictions and developing custom processing workflows.
- **Use advanced surrogates or other techniques to better represent pedestrians and bicyclists.** The surrogate pedestrians and bicyclists were necessary but could not mimic the full range of natural human movement patterns, impacting the realism of detection and prediction tasks.
- **Plan robust data storage, distribution, and backend infrastructure.** Stage 1B required robust backend infrastructure to support data collection, sharing, and submissions.

Thermal image enhancement example for visual camera-thermal camera pairwise extrinsic calibration.

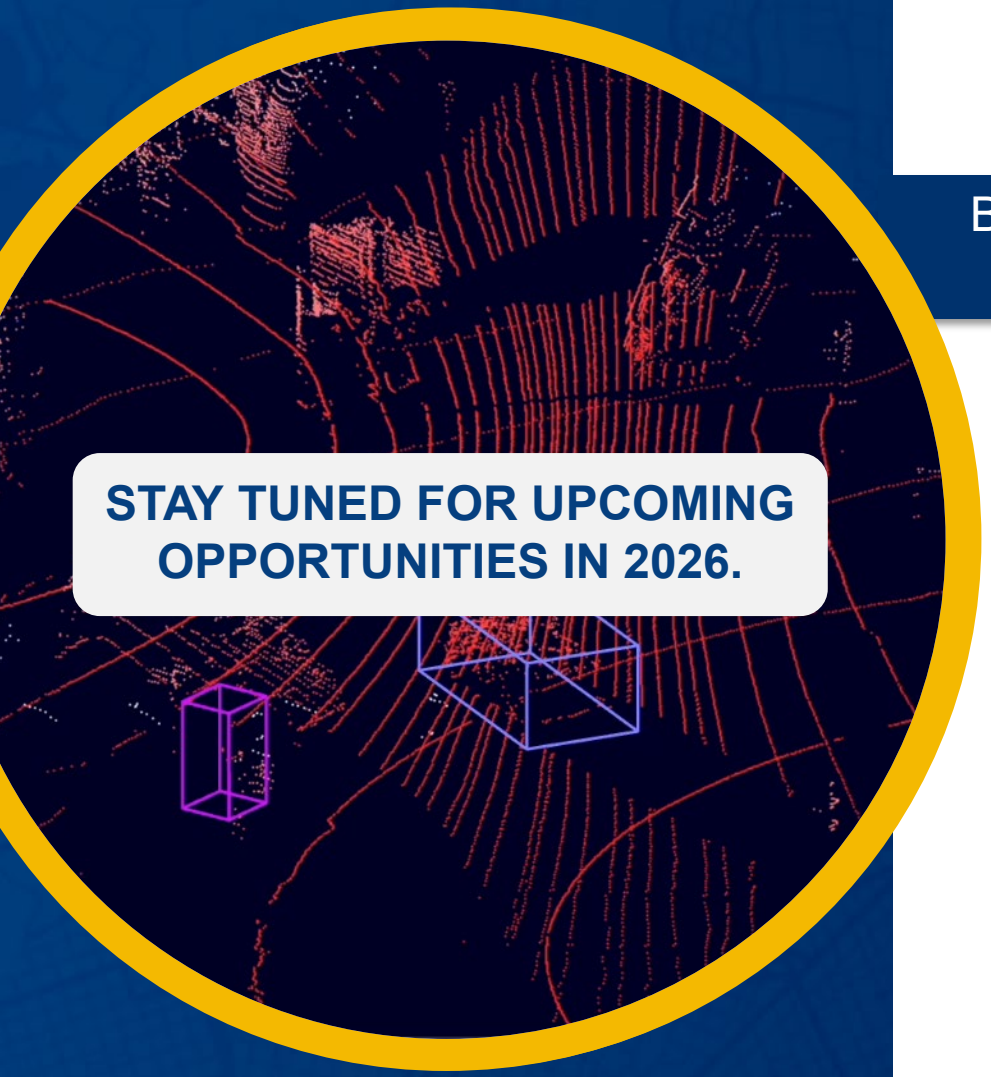
Source: FHWA

Future Direction

Intersection Safety Systems (ISS)

Norah Ocel, ITS JPO

Image Source: U.S. DOT



STAY TUNED FOR UPCOMING OPPORTUNITIES IN 2026.

Image Source: U.S. DOT

U.S. Department of Transportation
ITS Joint Program Office

INTERSECTION SAFETY SYSTEMS

Building on the success of the U.S. DOT Intersection Safety Challenge, this initiative aims to advance end-to-end system prototyping.

SENSE

Sense the intersection conditions using emerging, low-cost sensors (e.g., visual camera, LiDAR, radar, thermal camera) and data fusion.

THINK

Use artificial intelligence (AI) to improve situational awareness and predict safety threats.

ACT

Mitigate unsafe conditions involving vehicles and other road users by issuing warnings and/or modifying signalized control settings.

Sources Sought Notice (May 2025)

Includes *Draft* Scope of Work for ISS Prototyping

- **Objective:** acquire ISS prototypes to assess the potential benefit and feasibility of broader ISS deployment to address high-value real-world intersection safety issues.
- **Note:** any contractor team must have both: a public sector partner that is interested in piloting an ISS solution, and access to a configurable controlled environment test bed
- **Key Research Questions:**
  - *Conflicts.* Can an ISS reliably predict conflicts among vehicles and vulnerable road users?
  - *Unsafe Conditions.* Can an ISS reliably identify conflicts and characterize unsafe conditions?
  - *Mitigating Strategies.* Can an ISS take meaningful mitigating actions, either in the form of warnings or changes to intersection control, to mitigate conflicts and/or unsafe conditions?

Questions?

Contact Information

Norah Ocel, P.E., PMP

Program Manager, Strategic Technology for Roadway Safety

Intelligent Transportation Systems Joint Program Office (ITS JPO)

Norah.Ocel@dot.gov