CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Final Report 118
July 2026

Prototyping a Low-cost Roadside Device System for Cooperative Automated Driving

Yang Cheng, PhD
Steven T. Parker, PhD
Bin Ran, PhD
Sicheng Fu, PhD
Jiaxi Liu
Rui Gan
Junwei You, PhD

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Final Report 118 *July 2026*

Project Start Date: 07/01/2024

Project End Date: 06/30/2025

Prototyping a Low-cost Roadside Device System for Cooperative Automated Driving

Integrated CRM/UWB Localization, Work Zone Geometry Reconstruction, and WZDx-Ready Navigation Refinement

Yang Cheng, cheng8@wisc.edu

Project Investigator, Research Scientist

Bin Ran, bran@wisc.edu

Vilas Distinguished Achievement Professor of Civil and Environmental Engineering and Director of the ITS Program

Steven T. Parker, sparker@engr.wisc.edu

Managing Director of the Wisconsin Traffic Operations and Safety (TOPS) Laboratory

The University of Wisconsin - Madison

DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3552348305 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation:

Cheng, Y., Ran, B., Parker, S. T., Fu, S., Liu, J., Gan, R., You, J. (2026). Prototyping a Low-cost Roadside Device System for Cooperative Automated Driving: Integrated CRM/UWB Localization, Work Zone Geometry Reconstruction, and WZDx-Ready Navigation Refinement. Center for Connected and Automated Transportation, University of Wisconsin-Madison, Madison, WI.

Contacts

For more information:

Steven Parker

445 Henry Mall, Room 203

Madison, WI 53706

608-265-4921

sparker@engr.wisc.edu

CCAT

University of Michigan Transportation Research Institute

2901 Baxter Road

Ann Arbor, MI 48152

umtri-ccat@umich.edu

(734) 763-2498

Technical Report Documentation Page

1. Report No. Final Report 118	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Prototyping a Low-cost Roadside Device System for Cooperative Automated Driving: Integrated CRM/UWB Localization, Work Zone Geometry Reconstruction, and WZDx-Ready Navigation Refinement		**5. Report Date** July 2026 **6. Performing Organization Code** N/A
7. Author(s) Yang Cheng, Ph.D., https://orcid.org/0009-0007-6102-4742 Steven T. Parker, Ph.D., https://orcid.org/0000-0003-3104-5116 Bin Ran, Ph.D., https://orcid.org/0000-0002-3269-6909 Sicheng Fu, Ph.D., https://orcid.org/0000-0002-5918-8891 Jiaxi Liu, https://orcid.org/0009-0001-2749-6435 Rui Gan, https://orcid.org/0009-0007-1344-9703 Junwei You, Ph.D., https://orcid.org/0009-0002-6447-8276		**8. Performing Organization Report No.** **N/A**
9. Performing Organization Name and Address Center for Connected and Automated Transportation University of Michigan Transportation Research Institute 2901 Baxter Road Ann Arbor, MI 48109 Department of Civil and Environmental Engineering University of Wisconsin-Madison 1415 Engineering Drive Madison, WI 53706		**10. Work Unit No.** **11. Contract or Grant No.** Contract No. 69A3552348305

12. Sponsoring Agency Name and Address	**13. Type of Report and Period Covered**
U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE Washington, DC 20590	Final Report (July 2024 - June 2026) **14. Sponsoring Agency Code** OST-R

15. Supplementary Notes

Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract

Although significant progress has been made in automated driving technologies, technical challenges still exist, especially for complex Operational Design Domains (ODDs). A low-cost roadside device system, the Connected Reference Marker (CRM) System, has been developed to support CAVs in those ODDs. The CRM system can facilitate CAV localization by providing real-time distance measurement and road geometry changes (i.e., work zones). Therefore, the CRM system has the potential to serve as a gateway system for infrastructure-based cooperative driving automation (CDA) due to its low cost and easy deployment. This project will evaluate the performance regarding localization and road geometry data provision in field experiments. Specifically, this project will build a prototype system and evaluate the localization accuracy in various scenarios; in addition, the prototype system will be used to detect the boundaries of work zones, as improving access to work zone data is one of the top needs identified through the USDOT Data for Automated Vehicle Integration (DAVI) effort. The detected boundaries of work zones will be later translated into a data feed following the Work Zone Data Exchange (WZDx) specification, a national work zone data standard pioneered by USDOT to meet the DAVI requirement and a critical part of the Roadway Digital Infrastructure (RDI) strategy.

17. Key Words	**18. Distribution Statement**
Connected Reference Marker; Ultra-Wideband; Cooperative Driving Automation; Work Zone Data Exchange; CAV localization; Work zone geometry; Large Language Models	No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	40	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

Executive Summary

1. Introduction

2. Project Description and Proposal Alignment

3. Problem Setup

4. Proposed Pose-Conditioned Permutation-Equivariant Predictive Denoiser

5. Experimental Setup

6. Findings

7. CRM System Prototype and Baseline Localization Evaluation

8. CRM-Enhanced WZDx and LLM-Based Work Zone Navigation Refinement

9. Integrated Technical Conclusions

10. Recommendations

11. Final Documentation of Outputs, Outcomes, and Impacts

12. Experimental Dataset Documentation and Readiness Package

13. Challenges and Lessons Learned

14. Acknowledgments

15. References

Appendix A. Performance Indicator Synopsis

Appendix B. Proposal Task Compliance Matrix

Appendix C. Experimental Dataset Data Dictionary and Acceptance Checklist

List of Figures

Figure 1. Work zone UWB ranging scenario and the two dominant ranging error modes: burst outliers and NLOS bias.

Figure 2. Overall pipeline of the V2I-enabled inference backbone and the two-stage denoising workflow.

Figure 3. Real-world V2I experiment setup: cone-mounted UWB-RSU, in-vehicle UWB tag/RTK-GNSS mounting, heights, anchor layout, and driven trajectory.

Figure 4. Statistical overview of the clean and challenging simulation scenarios.

Figure 5. Progressive work zone geometry reconstruction on one real dynamic episode.

Figure 6. Time-series comparison on one anchor stream from the challenging scenario; shaded regions indicate NLOS-active segments.

Figure 7. Absolute-error distribution on the challenging scenario, shown as CDF and complementary CDF.

Figure 8. Anchor-dropout robustness on the challenging scenario.

Figure 9. Downstream robustness under anchor perturbations on the challenging scenario.

Figure 10. Pose-noise sensitivity of pose-conditioned denoisers on the challenging scenario.

Figure 11. UWB two-way ranging process used by the CRM distance-measurement module.

Figure 12. Four-anchor CRM/Tag architecture for advanced triangular positioning.

Figure 13. Four-anchor triangular positioning geometry and coordinate averaging.

Figure 14. Error distribution of CRM combinations in the static scenario.

Figure 15. Static-scenario two-dimensional positioning clusters and ground-truth comparison.

Figure 16. Error distribution of CRM combinations in the dynamic scenario.

Figure 17. Estimated moving trace in Dynamic Experiment 1.

Figure 18. Estimated moving trace in Dynamic Experiment 2.

Figure 19. CRM-enhanced refinement of a WZDx right-lane closure: overview map.

Figure 20. CRM-enhanced refinement of a WZDx right-lane closure: zoomed map.

Figure 21. CRM-enhanced refinement of a curved multi-lane closure: overview map.

Figure 22. CRM-enhanced refinement of a curved multi-lane closure: zoomed map.

List of Tables

Table 1. Abbreviations used in this report.

Table 2. Proposal task-to-final-report alignment matrix.

Table 3. Real-world dynamic-episode results. Signal-level range error is weighted by valid measurement count; downstream geometry uses shared solvers. Lower is better except IoU.

Table 4. Range denoising MSE on challenging and clean test splits. Lower is better.

Table 5. Base versus polygon-tuned checkpoints of the proposed method on the challenging scenario test split.

Table 6. Bootstrap robustness for episode-weighted overall MSE, reported as mean episode-level MSE [95% CI].

Table 7. Tail percentiles and burst-segment metrics on the challenging scenario test split. Lower is better.

Table 8. Anchor-stress robustness in the challenging scenario: overall MSE under anchor dropout and random episode-consistent permutations.

Table 9. Cone-position accuracy and work zone shape stability under 20 random episode-consistent anchor permutations on the challenging scenario test split.

Table 10. Downstream robustness under anchor dropout on the challenging scenario. Polygon IoU is averaged over non-degenerate episodes.

Table 11. Component ablation on the challenging scenario test split.

Table 12. Pose-noise sensitivity on the challenging scenario when zero-mean planar noise is injected into the model pose input at test time.

Table 13. Output documentation by required reporting category.

Table 14. Outcome and impact documentation.

Table 15. Experimental dataset acceptance-ready package contents.

Table 16. Challenges and lessons learned.

Table 17. Performance indicator synopsis.

Table 18. Proposal task compliance matrix.

Table 19. Dataset data dictionary.

Abbreviations

Table 1. Abbreviations used in this report.

Abbreviation	Definition
ATMS	Advanced Traffic Management System
CAV	Connected and Automated Vehicle
CDA	Cooperative Driving Automation
CRM	Connected Reference Marker
FOV	Field of View
GNSS	Global Navigation Satellite System
IoU	Intersection over Union
LOS/NLOS	Line-of-Sight / Non-Line-of-Sight
MSE/MAE	Mean Squared Error / Mean Absolute Error
OBU	Onboard Unit
ODD	Operational Design Domain
RSU	Roadside Unit
RTK/INS	Real-Time Kinematic / Inertial Navigation System
UWB	Ultra-Wideband
V2I	Vehicle-to-Infrastructure
WZDx	Work Zone Data Exchange
ADS	Automated Driving System
CADS	Collaborative Automated Driving System
CWZ	Connected Work Zones
DAVI	Data for Automated Vehicle Integration
LLM	Large Language Model
ToF	Time of Flight
WZDx	Work Zone Data Exchange

Executive Summary

This integrated final report documents the Connected Reference Marker (CRM) / ultra-wideband (UWB) cooperative automated driving project across three connected research outputs: (1) a baseline CRM hardware prototype and static/dynamic localization evaluation, (2) a V2I work zone geometry-reconstruction manuscript with pose-conditioned UWB range denoising, and (3) a CRM-enhanced WZDx/LLM workflow for work zone data provision and AV-ready navigation refinement.

Together, the work evaluates a low-cost infrastructure-based CDA concept in which temporary work zone markers act as connected reference points. The CRM devices provide direct UWB distance measurements; the denoising and mapping workflow converts noisy multi-anchor ranges into cone and polygon geometry; and the WZDx/LLM workflow translates CRM-derived cone coordinates and WZDx semantic lane-closure data into navigation instructions, recommended lanes, and refined centerline polylines.

The technical findings support the original proposal direction. The prototype experiments show that CRM/UWB can achieve low static and dynamic localization errors under controlled conditions, while also revealing Y-axis and motion-related sensitivities that require calibration. The denoising study shows that pose-conditioned, permutation-equivariant modeling is most valuable in challenging NLOS and heavy-tailed regimes; in measured field data, the calibrated proposed method reduces signal-level MSE by 66.9% relative to calibrated raw input and ranks first on field anchor error and polygon IoU among the tested pipelines. The WZDx/LLM study demonstrates that precise CRM cone coordinates can correct coarse or misaligned WZDx polylines and produce interpretable, geometry-consistent work zone guidance.

The revised outputs, outcomes, and impacts section is written toward CCAT final-report categories and the 2024 proposal commitments: CRM system prototype, experimental datasets, project report, publications/manuscripts, analytical models, equipment documentation, technology-transfer readiness, and broader safety/mobility impacts. The dataset section now covers CRM static/dynamic data, V2I UWB denoising data, WZDx/LLM case-study inputs and outputs, data dictionary, manifest, quality-control checklist, and reproducibility notes. Raw data files have been submitted to Dryad with DOI https://doi.org/10.5061/dryad.xgxd254xf.

1. Introduction

Reliable work zone mapping is important for connected and autonomous vehicles (CAVs) and advanced traffic management systems because work zone geometry and traffic-control devices can change rapidly. Lane shifts, tapers, narrowed shoulders, cone lines, and temporary barriers may be installed, moved, or removed within hours or days, posing challenges for CAVs with uncertainties (H. Li et al. 2025). As a result, CAVs require an up-to-date representation of the active work area and the remaining drivable region to navigate safely and smoothly through work zones (Dehman and Farooq 2021). When visibility is limited or when the layout changes faster than onboard perception can update, vehicle-only sensing may provide an incomplete picture. Vehicle-to-infrastructure (V2I) sensing can help address this limitation by allowing roadside units to provide direct geometric information to passing vehicles (Liu et al. 2024). Recent work zone data initiatives, such as the Work Zone Data Exchange (WZDx), further highlight the need for harmonized and machine-readable work zone information that can be consumed by navigation systems, original equipment manufacturers, and automated driving systems (U.S. Department of Transportation 2024). However, static work zone data feeds do not provide dynamic map descriptions observed by a passing vehicle, which motivates real-time cone localization and precise work zone geometry reconstruction.

Existing work zone mapping pipelines mainly rely on three types of information. First, survey-grade or infrastructure-based mobile mapping systems can recover detailed roadway geometry, such as lane-width and cross-section changes, from calibrated sensors (e.g., LiDAR) (Habib et al. 2018). Second, vision-based approaches detect temporary traffic control devices (TTCDs) in work zones, such as cones, barrels, signs, workers, and equipment, and then infer work zone extent through scene understanding and topology-based reasoning (Seo et al. 2022; Zuo et al. 2023; Shi and Rajkumar 2021). Third, trajectory-based approaches use crowd-sourced vehicle trajectories to infer temporarily changed drivable areas and support downstream navigation (Chen, Luo, and Feng 2023). These approaches are valuable but have complementary limitations. Mapping-grade systems provide accurate geometry but require dedicated equipment and calibration. Vision-based methods are sensitive to occlusion, lighting, and object appearance variation, as well as limited by annotation effort, remaining a long-tailed problem (Ghosh et al. 2025). Trajectory-based methods may respond slowly to abrupt layout changes or become unreliable when vehicle observations are sparse. A temporary work zone sensing system should therefore be low-cost, rapidly deployable, less dependent on visual scene understanding alone, and able to provide direct metric constraints as the layout changes.

Ultra-wideband (UWB) ranging is promising for this role because it provides direct distance measurements with fine time resolution and can be deployed with low-cost roadside hardware (Gezici et al. 2005; Alarifi et al. 2016; Volpi et al. 2023). UWB has also been used for tracking and hazard zone monitoring in construction and roadwork safety applications (Maalek and Sadeghpour 2016; Ochoa-de-Eribe-Landaberea, Zamora-Cadenas, and Velez 2024). In a V2I work zone mapping system, UWB-enabled roadside units (UWB-RSUs) can be mounted on traffic cones and treated as temporary roadside anchors. A vehicle-mounted UWB tag then measures ranges to these anchors while the vehicle passes through or near the work zone, turning the cone layout into a set of constraints for cone localization and work zone geometry reconstruction. In this way, UWB does not replace survey-grade, vision-based, or trajectory-based mapping, but provides a complementary low-cost ranging layer for temporary work zone deployments.

The remaining challenge is that field UWB ranges must be reliable enough to support this downstream mapping task. Non-line-of-sight (NLOS) propagation and multipath can introduce positive range bias, while transient interference can create burst outliers and heavy-tailed residuals (F. Wang, Tang, and Chen 2023; Angarano et al. 2021). Classical filtering and robust state-space methods, including adaptive Kalman filtering for time-of-flight systems, provide interpretable and low-latency baselines (Q. Li et al. 2015). However, simple per-step filters may be less reliable when NLOS errors persist over multiple time steps or when large impulses appear in short bursts. Learning-based UWB error mitigation can capture richer nonlinear corrections, especially when physical-layer signals or temporal context are available (Angarano et al. 2021; Niu et al. 2023; Yang et al. 2024). At the same time, purely supervised learning remains challenging because accurate ground-truth distances require survey-grade references that are expensive to collect at scale. Semi-supervised and predictive learning strategies therefore offer a useful direction by leveraging abundant unlabeled range streams before supervised calibration (T. Wang et al. 2021; Y. Li, Mazuelas, and Shen 2023; Assran et al. 2023). These limitations motivate a denoising method that first learns temporal regularity from observed range sequences and then uses available ground-truth distances for supervised refinement.

In addition to range noise, a V2I work zone mapping system must handle changes in the anchor set. The number of visible UWB-RSUs can vary across deployments and across time because of occlusion, equipment movement, temporary blockage, or operational changes. Moreover, raw ranging interfaces may return visible anchors as unordered anchor-range pairs, so the same physical anchors can appear under different input orders after software-level re-indexing. A fixed-order vector representation can therefore become fragile when the input order changes or when some anchors are missing. Set-structured learning provides a natural way to address this issue through shared element-wise transformations and symmetric aggregation, as formalized in Deep Sets and extended in attention-based set models (Zaheer et al. 2017; Lee et al. 2019). For work zone UWB ranging, this means that the denoiser should treat the anchor dimension as a masked set rather than as a fixed-order vector. Robustness to anchor re-indexing and anchor dropout is therefore a practical deployment requirement rather than only a modeling preference.

To address these challenges, this paper studies UWB-assisted work zone mapping as a V2I geometry-inference problem. The central idea is to improve the reliability of multi-anchor UWB ranges before they are used for cone localization and work zone geometry reconstruction. We propose a pose-conditioned, permutation-equivariant predictive denoiser for multi-anchor UWB ranging. The model combines shared anchor-wise temporal prediction to capture range dynamics, symmetric set aggregation to handle unordered and missing anchors, and pose-conditioned residual decoding to use the vehicle pose stream as a geometric prior. The model is trained in two stages: predictive pretraining on observed ranges followed by NLOS-weighted supervised fine-tuning. This design aims to improve range accuracy under bursty NLOS errors while maintaining consistency under anchor re-indexing and dropout.

Our contributions are threefold:

We formulate UWB-assisted work zone mapping as a V2I geometry-inference problem, in which cone-mounted UWB-RSUs serve as anchors and multi-anchor range measurements provide direct constraints for cone localization and work zone geometry reconstruction.

We develop a pose-conditioned, permutation-equivariant predictive denoiser for multi-anchor UWB ranging. The model combines anchor-wise temporal prediction, symmetric set aggregation, and pose-conditioned residual decoding to cope with burst outliers and NLOS errors while handling unordered anchor inputs and missing anchors.

We validate the proposed method on both measured V2I field data and large-scale controlled simulations, evaluating range denoising, cone localization, work zone geometry reconstruction, anchor re-indexing and dropout robustness, component sensitivity, and pose-noise sensitivity.

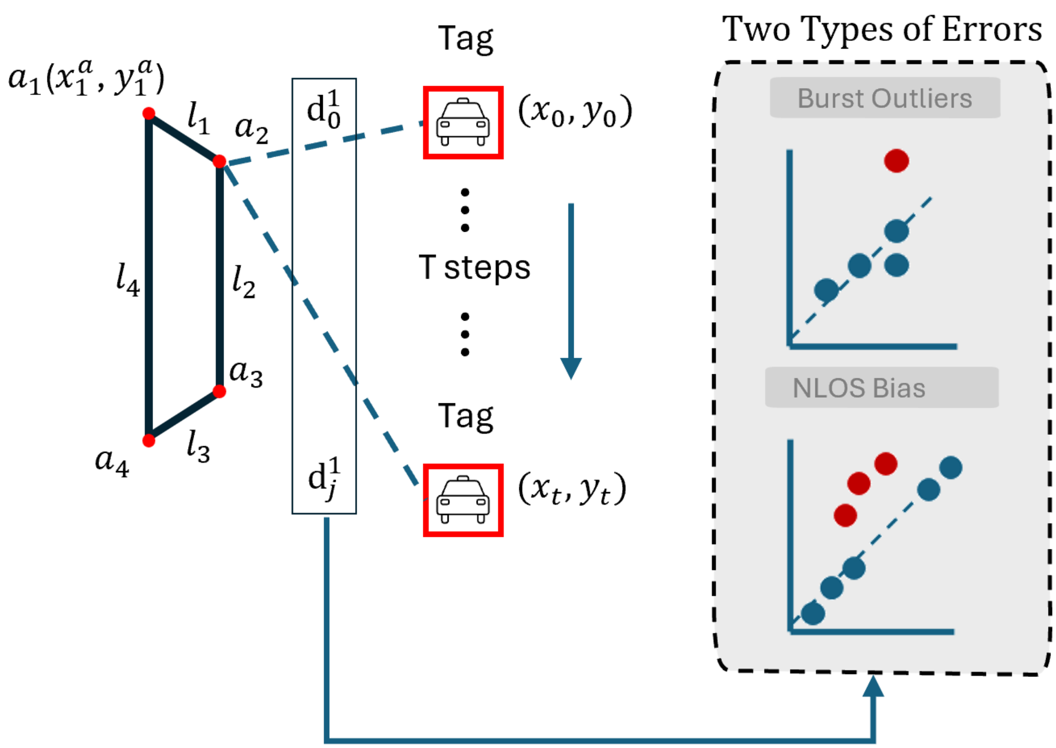

Figure 1. Work zone UWB ranging scenario and the two dominant ranging error modes: burst outliers and NLOS bias.

2. Project Description and Proposal Alignment

The original proposal describes the Connected Reference Marker (CRM) system as a low-cost roadside device system developed to support CAV localization through real-time distance measurement and to provide detailed road geometry, including work zone geometry, for safe CAV navigation. The final report therefore keeps the technical manuscript centered on UWB-based ranging, work zone boundary inference, field evaluation, simulation benchmarking, and downstream geometry reconstruction.

The proposal identifies three core CRM components: roadside devices, a cloud service, and vehicle onboard units. The roadside side includes a CRM array, server, and RSU; the vehicle side includes an OBU with a CRM/UWB unit. The practical work zone use case installs anchor CRMs and tag CRMs on cones or temporary traffic control devices, sends raw position and range data to a local server, computes the work zone geographic boundary, and broadcasts or uploads the resulting information for use by CAVs, map providers, DOT ATMS, and lane-closure systems.

Table 2. Proposal task-to-final-report alignment matrix.

Proposal task	Original intent	Final report coverage
Task 1. System Assembly	CRM/UWB hardware and software assembled for roadside/vehicle-side experiments.	Covered by system description, field setup, and equipment/output sections.
Task 2. Distance Measurement Calibration	Calibration and evaluation of CRM distance measurement under operating conditions.	Covered by UWB observation model, field calibration, range denoising metrics, and dataset QA.
Task 3. Vehicle Localization Performance Evaluation	Evaluation of CRM localization performance for vehicles across scenarios.	Covered by tag trajectory, cone localization, Anchor Error_w, and downstream geometry analysis.
Task 4. Work Zone Boundary Mapping with CRM Integration	Work zone boundary mapping using CRM-provided distances and GPS/pose data.	Covered by work zone polygon reconstruction, IoU, Hausdorff distance, and progressive reconstruction figures.
Task 5. WZDx Data Feed	Prototype CRM system as a data source for WZDx feed and data dictionary mapping.	Covered by outputs/outcomes and dataset sections; WZDx mapping is documented as a recommended deployment pathway.
Task 6. Project Report and Technical Transfer	Prepare final report, presentations, and publications.	Covered by this revised final report, manuscript citation, performance indicators, and technology transfer outputs.

This alignment section is included so that the final report can be evaluated against both the paper content and the proposal commitments. The later output/outcome/impact and dataset sections expand the acceptance-oriented documentation beyond the research manuscript itself.

3. Problem Setup

V2I Sensing Scenario, Data Representation, and Task Definition

We consider a V2I-enabled sensing system in which UWB-RSUs provide infrastructure-side ranges from work zone cones to a moving vehicle-mounted tag. Each UWB-RSU is a UWB anchor mounted on a traffic cone and treated as a single fixed infrastructure-side ranging anchor. Figure 1 shows the sensing geometry together with the two error patterns that motivate the formulation below. Burst outliers are isolated deviations caused by transient interference or abrupt multipath events. NLOS bias is a sustained positive shift that appears when the direct ranging path is blocked and reflections dominate. These two effects have distinct temporal signatures, so the method later combines temporal prediction with bias-aware method rather than treating all errors as independent, identically distributed noise.

A work zone episode is a trajectory segment of length T in which the vehicle-mounted tag measures ranges to a set of UWB-RSUs. We index episodes by $e \in \{1, \dots, E\}$, time steps by $t \in \{1, \dots, T\}$, and anchor slots by $i \in \{1, \dots, N_{\max}\}$; each episode is right-padded to $N_{\max}$ for batch processing, and the structural mask separates real anchor slots from padded ones. We write $\mathcal{T} = \{1, \dots, T\}$ and $\mathcal{N} = \{1, \dots, N_{\max}\}$ for the time and padded anchor-slot index sets. Anchor identity is fixed within an episode in our benchmarks, and sustained anchor unavailability is evaluated separately through anchor-dropout stress tests.

For a fixed episode and indices t, i, the raw inputs are the vehicle-mounted tag position $p_t \in \mathbb{R}^2$, the observed raw range $d_{t,i}^{obs}$, and the structural validity mask $m_{t,i} \in \{0,1\}$.

We additionally have ground-truth ranges $d_{t,i}^{gt}$ and an NLOS indicator $n_{t,i} \in \{0,1\}$ for supervised training and diagnostics during training phase. Collecting these objects at the episode level gives:

$$\boldsymbol{P} = \{p_t\}_{t\in\mathcal{T}}, \quad \boldsymbol{D}^{obs} = \{d_{t,i}^{obs}\}_{t\in\mathcal{T},\, i\in\mathcal{N}}, \quad \boldsymbol{M} = \{m_{t,i}\}_{t\in\mathcal{T},\, i\in\mathcal{N}}.$$

The denoiser f_θ maps these sequence inputs to corrected ranges:

$$\hat{d}_{t,i} = f_\theta(\boldsymbol{P}, \boldsymbol{D}^{obs}, \boldsymbol{M})_{t,i}.$$

The goal in this range phase is to minimize overall, line-of-sight (LOS), and NLOS range error for UWB-enabled distance estimation and also preserve prediction consistency under episode-consistent anchor permutation and anchor dropout situations.

UWB Observation Model and Error Structure

Let $a_i = (x_i^a, y_i^a) \in \mathbb{R}^2$ denote the 2D coordinate of the infrastructure anchor associated with slot i, and write the reference tag position as $p_t = (x_t, y_t) \in \mathbb{R}^2$. The ground-truth distance is:

$$d_{t,i}^{gt} = \| p_t - a_i \|_2.$$

Anchor identity and episode alignment.

Throughout, the anchor index i denotes an anchor identity that is consistent across time within an episode (e.g., a stationary UWB-RSU broadcasting a unique ID, or an anchor with a known fixed coordinate that can be matched by proximity). In practice, raw ranging APIs often return an unordered list of anchor id and raw measured distance pairs whose list order may vary across calls. a lightweight normalized step aligns these measurements into a fixed index order. Under this alignment, software-induced order changes are well modeled as an episode-consistent permutation, which Section 3.3 formalizes as the structural property the denoiser must satisfy.

The observed range is modeled as:

$$d_{t,i}^{obs} = m_{t,i}\big(d_{t,i}^{gt} + b_{t,i} + \beta_{t,i}^{+} + \varepsilon_{t,i}\big),$$

where $b_{t,i}$ denotes signed per-anchor offset and drift, $\beta_{t,i}^{+} \geq 0$ denotes NLOS inflation, and $\varepsilon_{t,i}$ collects zero-mean transient perturbations, including temporally correlated noise and occasional signed heavy-tailed impulses. This decomposition separates signed anchor-specific calibration error, sustained positive propagation inflation, and transient perturbations. In Figure 1, burst outliers mainly appear as large transient realizations of $\varepsilon_{t,i}$, whereas NLOS segments are characterized by sustained positive $\beta_{t,i}^{+}$ together with heavier-tailed residual variation. For simulation analysis, an indicator $n_{t,i}$ marks whether a sample is NLOS-affected. It is used for diagnostics and loss weighting.

The base LOS measurement component is modeled as zero-mean Gaussian noise, but the benchmark also adds per-anchor bias and drift and occasional heavy-tailed impulses, so LOS-labeled samples can still show non-Gaussian residual tails. In the NLOS state ($n_{t,i} = 1$), additional positive, time-varying bias terms become active. Section 4.2 instantiates these states with concrete parameters for different simulation regimes.

The denoiser takes p_t as an auxiliary input to leverage motion regularity and geometry context. In deployment, p_t will come from RTK/INS. To avoid unrealistic information leakage, we later evaluate injected pose perturbations at test time rather than assuming perfect pose.

4. Proposed Pose-Conditioned Permutation-Equivariant Predictive Denoiser

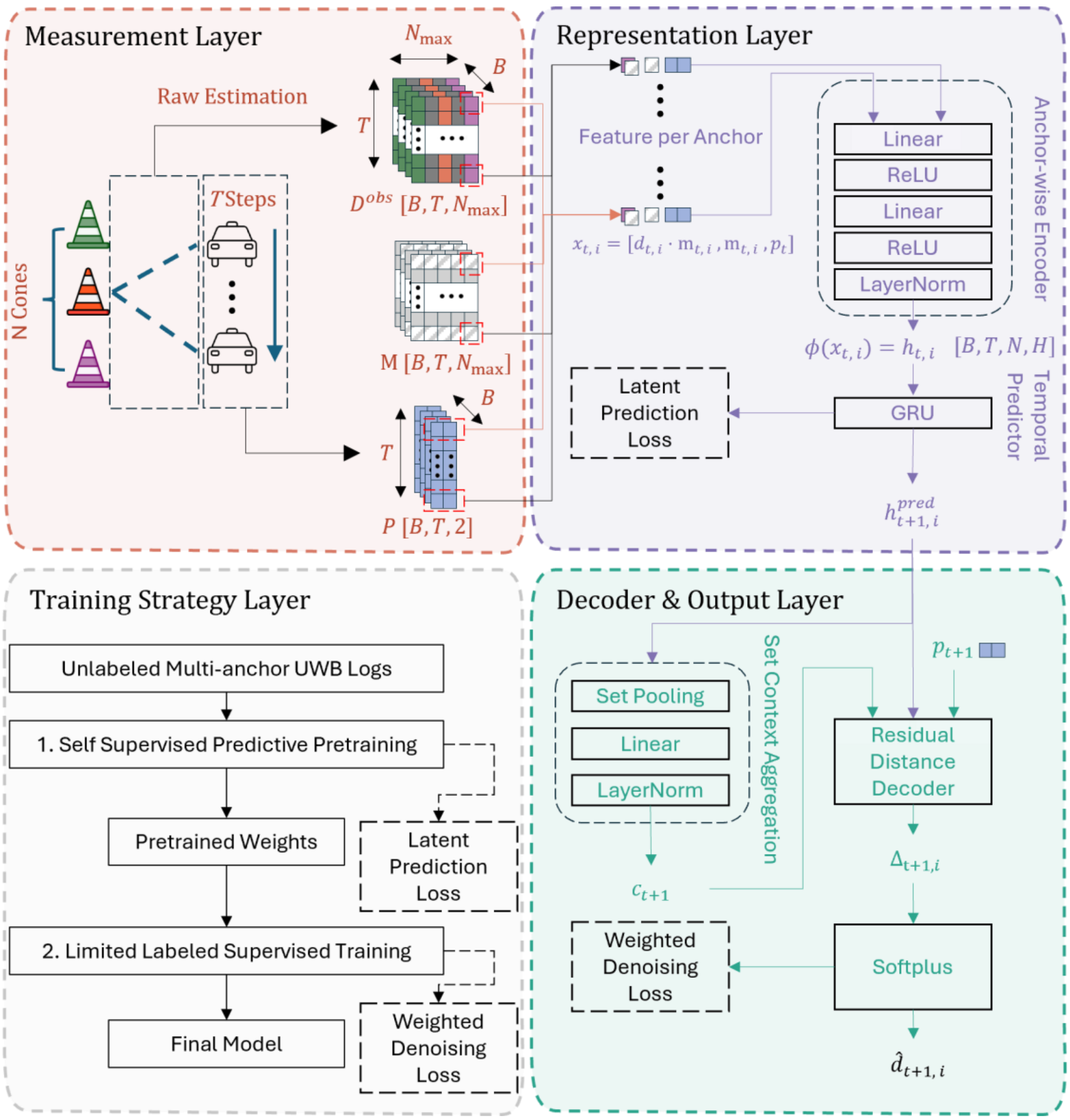

Figure 2. Overall pipeline of the V2I-enabled inference backbone and the two-stage denoising workflow.

Figure 2 shows the proposed pipeline. The architecture combines four components: shared anchor-wise temporal modeling for bursty range errors as shown in the Representation Layer, symmetric set-level aggregation for unordered and variable-cardinality anchor observations as shown in the Decoder Output Layer, pose-conditioned residual decoding for geometry-aware correction as shown in the Decoder Output Layer, and a two-stage training protocol with predictive pretraining followed by NLOS-aware supervised fine-tuning as shown in the Training strategy layer. The remainder of this section specifies architecture of the backbone, training objectives, equivariance property, and inference behavior.

Architecture of the Predictive Denoiser

The model takes $\boldsymbol{P}$, $\boldsymbol{D}^{obs}$, and $\boldsymbol{M}$ as inputs. At each step it first forms a predictive latent prior for time $t+1$ from the history up to t, then combines this prior with the same-step raw measurement through residual decoding. The backbone outputs $\hat{d}_{t+1,i}$ together with predictive latents $\hat{h}_{t+1,i}$ used during training. When referring to the aligned denoised sequence generically, we write $\hat{d}_{t,i}$. The validity mask gates padded anchor slots in the forward pass, while the NLOS indicator $n_{t,i}$ is used only in supervised loss weighting during training and never as a denoiser input at inference time.

Step 1: Local anchor-state encoding. Let $\phi(\cdot)$ denote a shared multilayer perceptron (MLP) applied independently to each anchor stream. We build local features as:

$$h_{t,i} = \phi\left([d_{t,i}^{obs} m_{t,i},\ m_{t,i},\ p_t]\right).$$

Including p_t in the per-anchor input allows the encoder to condition each range stream on the local vehicle state from the first layer.

Step 2: Shared temporal prediction. Let GRU($\cdot$) denote the shared gated recurrent unit (GRU) used as the temporal predictor. Anchor-wise temporal prediction is performed by:

$$\hat{h}_{t+1,i} = \mathrm{GRU}(h_{1:t,i}).$$

Using one shared predictor across anchors encourages consistent dynamics modeling and reduces anchor-ID-specific overfitting. By accumulating recurrent state along each anchor's history, this step captures temporal structure that is difficult for per-step filters or memoryless feed-forward learners to exploit under bursty, non-Gaussian range errors.

Step 3: Set-level context aggregation. Let $\rho(\cdot)$ denote a shared context MLP, and let $\varepsilon_0 > 0$ denote a small numerical-stability constant. Set-level context at each step is pooled from predicted latents:

$$c_{t+1} = \rho\left(\frac{\sum_i m_{t+1,i}\, \hat{h}_{t+1,i}}{\sum_i m_{t+1,i} + \varepsilon_0}\right).$$

This context provides cross-anchor coupling while remaining permutation invariant through symmetric averaging. Combined with the shared anchor-wise operators, this design yields equivariance to anchor re-indexing and provides a mask-aware interface for the dropout stress tests.

Step 4: Residual distance decoding. Let $g(\cdot)$ denote the shared residual decoder. Residual decoding predicts a correction to the same-step raw range:

$$\begin{aligned} \Delta_{t+1,i} &= g([\hat{h}_{t+1,i}, c_{t+1}, p_{t+1}]), \\ \hat{d}_{t+1,i} &= \mathrm{softplus}(d_{t+1,i}^{obs} + \Delta_{t+1,i}) m_{t+1,i}. \end{aligned}$$

The decoder uses the predictive latent state and the pose-aware set context as priors, correcting the same-step observation rather than regressing an absolute distance from scratch. The softplus nonlinearity enforces non-negative range outputs. Pose is therefore used both in anchor-wise encoding and residual decoding, allowing the network to condition corrections on local motion and global geometry.

Training Objectives and Two-Stage Learning

Stage 1 (unlabeled predictive pretraining). Let sg($\cdot$) denote the stop-gradient operator, which blocks gradient flow through its argument so that the predictor is trained toward a fixed latent target. The first phase uses unlabeled range logs to learn temporal regularity and cross-anchor latent structure:

$$\mathcal{L}_{pred} = \mathrm{MaskedMSE}\left(\hat{h}_{t+1,i}, \mathrm{sg}\left(h_{t+1,i}\right)\right),$$
$$\mathcal{L}_{dec}^{obs} = \mathrm{MaskedHuber}\left(\hat{d}_{t+1,i}, d_{t+1,i}^{obs}\right).$$

$$The Stage 1 objective is: \mathcal{L}_{stage1} = \mathcal{L}_{pred} + \lambda_{obs}\mathcal{L}_{dec}^{obs},$$

where $\lambda_{obs} \geq 0$ controls the strength of the observation-space auxiliary term. The masked losses are evaluated only over valid anchor slots, and the Huber penalty keeps fitting robust to heavy-tailed outliers, which is quadratic for small residuals, linear for large ones. This stage teaches the model to forecast temporally coherent latent evolution and to decode measurements without requiring ground-truth distance labels. The anchor-wise encoder, shared GRU predictor, set-level context module, and residual decoder are optimized jointly.

Stage 2 (supervised denoising fine-tuning). The second phase switches targets from observed range to ground truth:

$$w_{t+1,i}^{sup} = m_{t+1,i}\left[\left(1 - n_{t+1,i}\right) + \alpha n_{t+1,i}\right], \quad 0 < \alpha < 1,$$
$$\mathcal{L}_{sup} = \mathrm{WeightedHuber}\left(\hat{d}_{t+1,i}, d_{t+1,i}^{gt}; w_{t+1,i}^{sup}\right).$$

$$If predictive regularization is retained during fine-tuning, the Stage 2 objective becomes$$

$$: \mathcal{L}_{stage2} = \mathcal{L}_{sup} + \lambda_{pred}\mathcal{L}_{pred},$$

where $\lambda_{pred} \geq 0$ controls the extent to which the predictive prior is preserved after supervised calibration. Weighted supervision avoids over-penalizing difficult NLOS samples while still learning from them.

Design the role of the two stages.

The two stages play complementary roles in the architecture: Stage 1 supplies an initialization that already captures cross-anchor temporal regularity from raw observations, and Stage 2 then refines the same backbone toward the ground-truth distance target under NLOS-aware weighting. The empirical contribution of this Stage 1 initialization is quantified by the "w/o pretraining" row of the component ablation in Section 5.3.

Equivariance Property of the Proposed Architecture

We first state the structural requirement and then show the architecture satisfies it by construction. Let $S_{N_{\max}}$ denote the symmetric group on anchor slots, and let $\pi \in S_{N_{\max}}$ act on the padded range and mask tensors by $(\pi \boldsymbol{D})_{t,i} = \boldsymbol{D}_{t,\pi^{-1}(i)}$ and $(\pi \boldsymbol{M})_{t,i} = \boldsymbol{M}_{t,\pi^{-1}(i)}$. A denoiser f_θ that outputs anchor-wise corrected ranges is permutation equivariant if:

$$f_\theta(\boldsymbol{P}, \pi \boldsymbol{D}^{obs}, \pi \boldsymbol{M}) = \pi f_\theta(\boldsymbol{P}, \boldsymbol{D}^{obs}, \boldsymbol{M}), \quad \forall \pi \in S_{N_{\max}}.$$

This covers episode-consistent re-indexing, such as sorting by ID, by coordinate, or by an arbitrary API list order. The same masking interface lets anchor dropout be evaluated by zeroing selected anchor masks, but dropout robustness is assessed empirically rather than implied by equivariance alone. Time-varying permutations $\{\pi_t\}_{t=1}^{T}$, which represent data-association failures, are out of scope.

The proposed architecture satisfies f_θ by construction. The anchor-wise encoder $\phi(\cdot)$, temporal predictor $\mathrm{GRU}(\cdot)$, and residual decoder $g(\cdot)$ share parameters across anchors and are applied independently to each anchor stream, so re-indexing the inputs by π re-indexes their latent and output sequences in exactly the same way. The set-level context is a symmetric (masked-mean) aggregation and is therefore invariant to anchor order. Composing these steps gives the equivariance of f_θ for any $\pi \in S_{N_{\max}}$.

Inference Procedure and Computational Characteristics

At inference time only $(p_t, d_{t,i}^{obs}, m_{t,i})$ are needed. The model runs a single forward pass over the context window and feeds directly into the downstream solvers of Section 4.4. The dominant complexity is $\mathcal{O}(BTN_{\max}H^2)$ for the shared GRU plus $\mathcal{O}(BTN_{\max}H)$ for the shared MLP encode/decode, where B is batch size and H hidden width. Because the anchor operators are shared and independent before pooling, the architecture scales linearly in anchor count for fixed hidden width. These design choices make the model lightweight at inference time.

5. Experimental Setup

This section specifies how the formulation in Section 2 and the method in Section 3.1 are deployed in data, baselines, downstream solvers, and model-selection rules. Evaluation follows two complementary tracks: deployment-facing real-world V2I experiments and controlled large-scale simulation benchmarks.

Real-World V2I-Enabled Experiments

System setup and geometry instantiation.

The real dataset contains four infrastructure-side UWB-RSUs. Each UWB-RSU is mounted on a traffic cone and treated as one fixed anchor. In the current field dataset, the anchor layout is instantiated from the measured site dimensions as a trapezoidal work zone proxy with short base 2.69 m, long base 8.07 m, and height 5.13m, together with an explicit anchor-ID-to-corner assignment. The vehicle carries a UWB tag and an RTK/INS unit, with the global navigation satellite system (GNSS) antenna 1.55 m above ground. The UWB tag is mounted at 1.17 m above ground and offset from the GNSS antenna by 1.50 m forward and 0.56 m to the right. The reference tag trajectory is obtained by converting the RTK/INS latitude–longitude stream to a local planar frame and applying the measured rigid GPS-to-tag offset using GPS azimuth, so that the tag position used by the denoiser and the geometry-based range reference are spatially aligned. Because the UWB tag and the cone-mounted UWB-RSUs sit at different heights, the geometry-derived field reference distance is computed as the 3D slant range $d_{t,i}^{ref} = \sqrt{d_{xy,t,i}^2 + (h_{tag} - h_{anchor})^2}$, with $h_{tag} = 1.17$ m. The current field reference uses a ground-anchor approximation $h_{anchor} = 0$, reported explicitly as part of the calibrated field diagnostic. This slant reference matches the physical UWB range, whereas the downstream solvers of Section 4.4 use a 2D planar range model. The resulting tag-height frame difference applies identically to all methods and therefore does not affect the relative field comparison. Figure 3 summarizes the field setup, including the cone-mounted UWB-RSU hardware, the in-vehicle UWB-tag/RTK-GNSS mounting and lever arm, and the trapezoidal anchor layout with the driven trajectory overlaid on the site plan.

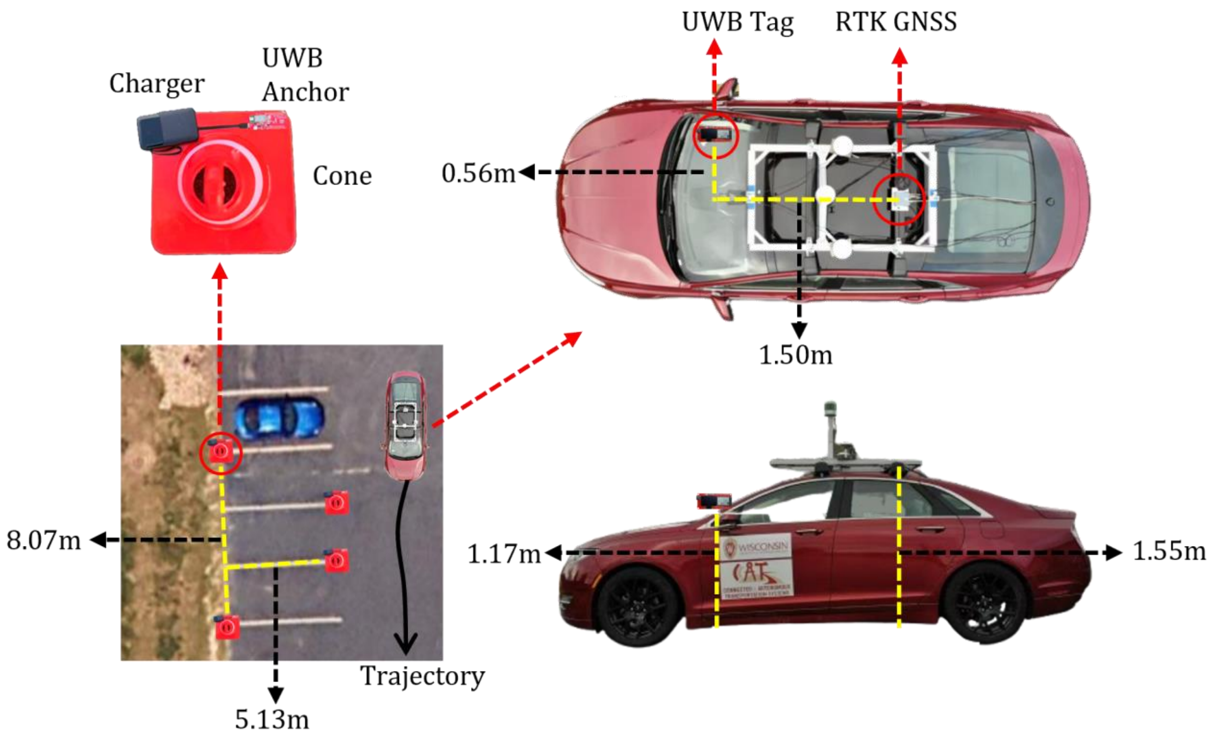

Figure 3. Real-world V2I experiment setup: cone-mounted UWB-RSU, in-vehicle UWB tag/RTK-GNSS mounting, heights, anchor layout, and driven trajectory.

Data collection and preprocessing.

Raw UWB events are aggregated into 0.10 s bins. Within each bin, the most recent measurement from each visible anchor is retained and missing anchors are left masked out. The dynamic reference trajectory is obtained from the RTK/INS stream after timestamp alignment and interpolation to the UWB event times. For each dynamic episode, a geometry-derived reference distance is then computed from the tag trajectory and the instantiated anchor layout. This yields a real-data sequence representation that matches the simulation-side interface: pose sequence, observed-range sequence, and validity mask sequence.

Evaluation protocol on real data.

The field track keeps the same family of denoisers as in simulation: raw input, Kalman (1D), MLP, PoseMLP, PoseKalman, and the proposed method as introduced in Section 4.3. The real-data analysis answers two questions. First, does the denoiser reduce signal-level ranging error relative to the geometry-derived reference distance? Second, do the denoised ranges remain useful for downstream localization and work zone geometry reconstruction under the same shared solvers used in simulation? The field reference distances come from measured site geometry plus RTK/INS-based tag pose, so the absolute downstream scale is bounded by the small site footprint. The relative ordering among the shared-input-calibrated pipelines is directly comparable. The proposed field-calibrated row is additionally reported as a calibrated diagnostic because it uses a degree-2 output calibration fitted to field reference distances.

Simulation Benchmarks for Controlled Large-Scale Evaluation

The simulation track instantiates the generic observation model of Section 2.2 in two regimes: the challenging scenario (burst NLOS, heavy-tailed outliers) with 1440 episodes and $T = 250$ steps, and the clean scenario (lower-noise, no burst corruptions) with 800 episodes and $T = 220$ steps. Both run at $50\,\text{Hz}$. Each episode samples an anchor count $N \sim \text{UnifInt}[4,12]$, anchor positions uniformly over a $60\,\text{m} \times 60\,\text{m}$ area, and a tag trajectory generated by a bounded random walk in speed and heading starting at $p_0 = (0,0)$. One anchor layout is reused across multiple trajectories. The observed range $d_{t,i}^{obs}$ adds six terms on top of the ground-truth distance: heteroscedastic (distance-dependent), first-order autoregressive (AR(1)) LOS noise, a per-anchor bias with drift, signed Laplace outliers, a positive NLOS-burst bias with geometric dwell time, a separate outage-burst corruption, and an offset for anchors outside the front-facing field of view (FOV). The challenging scenario uses all six terms and the clean scenario zeros out the two burst terms and uses a wider FOV, so the clean regime is a lower-noise negative control rather than an NLOS-free or purely Gaussian setting: per-anchor bias and drift, occasional outliers, and mild FOV-driven NLOS labels remain.

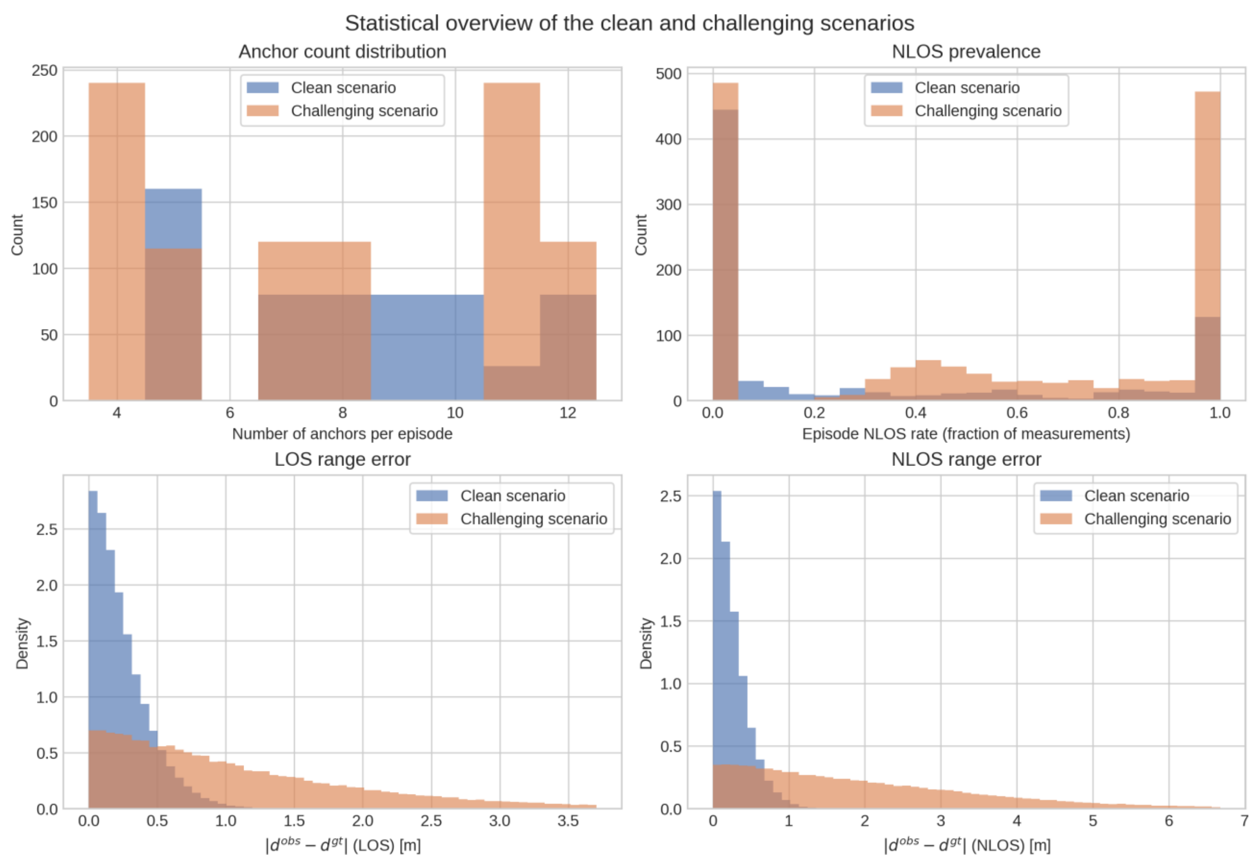

Figure 4. Statistical overview of the clean and challenging simulation scenarios.

Figure 4 confirms that the challenging scenario differs from the clean scenario along several axes: it contains more high-NLOS episodes and substantially heavier LOS/NLOS error tails while retaining a comparable anchor-count range. This supports using the challenging scenario as the primary stress-test.

Baselines and Comparison Protocol

We compare against four baselines:

independent per-anchor constant-position Kalman filter with scalar process noise Q and measurement noise R tuned on the validation split.

fixed-order per-step denoiser that consumes the padded multi-anchor range–mask vector without pose input.

fixed-order per-step denoiser that consumes the padded multi-anchor range–mask vector together with pose input.

adaptive Rauch–Tung–Striebel (RTS) smoother with pose-dependent process noise and NLOS measurement inflation. The inflation uses the simulation-only indicator n_t,i, so PoseKalman is an oracle-assisted baseline in the simulation benchmarks. On field data, where no label is available, it runs as the adaptive smoother without NLOS inflation.

All learned models use a strict episode-level 70/15/15 train/val/test split to avoid trajectory leakage. Metric masks (overall/LOS/NLOS), stress-test protocols, and downstream solvers are shared across methods, which keeps the comparison focused on the denoising backbone rather than protocol asymmetry. Checkpoint selection is validation-driven for every method and the per-method selection criterion is stated in the implementation details below.

Downstream Solvers and Evaluation Metrics

All methods feed their denoised ranges into the same downstream solver, so any downstream difference can be attributed to the input ranges. The deployment-relevant downstream task is cone localization under a known tag trajectory, which avoids the confounding of joint simultaneous localization and mapping (SLAM): the RTK/INS reference provides the tag positions $\{p_t\}$, and the denoised ranges $\hat{d}_{t,i}$ are used to estimate each cone (anchor) position $\hat{a}_i$ by a weighted nonlinear least-squares fit:

$$\hat{a}_i = \arg\min_{a\in\mathbb{R}^2} \sum_t w_{t,i}^{map} \left(\| p_t - a \|_2 - \hat{d}_{t,i}\right)^2,$$

solved by Gauss–Newton with a linearized least-squares (LS) initializer and damped refinement. The weights are $w_{t,i}^{map} = m_{t,i}[(1-n_{t,i}) + \alpha n_{t,i}]$ with $\alpha = 0.2$ when the oracle $n_{t,i}$ is available (simulation only), applied identically to all methods. These oracle-NLOS weights are a simulation-only diagnostic, and a deployment-facing pipeline would replace the term with an estimated reliability weight or simply use $w_{t,i}^{map} = m_{t,i}$. The recovered cone positions are then assembled into the work zone polygon, defined as the convex hull (the smallest convex polygon enclosing the mapped anchors) over the valid-anchor index set of an episode. For episodes that retain fewer than three valid mapped anchors, the convex hull is degenerate. Polygon intersection-over-union (IoU) and Hausdorff distance (the largest boundary-to-boundary gap between two polygons) are undefined for these episodes and are excluded from the reported polygon averages, which are taken over the non-degenerate episodes.

We report range MSE on LOS/NLOS/overall subsets together with tail percentiles ($p50$–$p99$), burst max-error and recovery duration, and anchor permutation/dropout stress-test MSE for signal-level quality. For downstream we report Anchor Error$_w$ (mean Euclidean error over valid anchors, measuring cone-position accuracy) in the main comparison, and polygon IoU plus symmetric Hausdorff distance d_H (measuring work zone geometry) in the model-selection and robustness analyses. Episode-level Spearman correlations between range MSE and downstream metrics, and degradation under injected planar pose noise, complete the diagnostic set. Polygon metrics can be non-monotonic in range MSE because the convex-hull proxy is sensitive to extreme anchors, which is why we report both signal-level and downstream metrics throughout.

Implementation Details and Model Selection

The final configuration of the proposed method uses hidden size 256, decoder hidden size 256, mean pooling, and context-based decoding. Stage 1, the model is pretrained on observed targets. Stage 2 then fine-tunes the model against ground-truth ranges with the weighted Huber loss on the full labeled training split. In the retained challenging-scenario checkpoint we set $\lambda_{pred} = 0$ during Stage 2, so fine-tuning optimizes only the supervised denoising loss once predictive pretraining has initialized the backbone.

The learned baselines (MLP and PoseMLP) retain the checkpoint with the lowest validation range loss. For the challenging scenario, the proposed method is additionally selected on the validation split by downstream polygon IoU under the fixed solver in Section 4.4. This polygon-tuned checkpoint is then reused throughout the main, robustness, and downstream analyses, and the checkpoint-selection asymmetry relative to the base checkpoint is analyzed in Table 1. The field evaluation uses the matched baseline implementations and reports the additional proposed-denoiser output calibration explicitly in the results section. All simulation results below are computed on held-out test data.

6. Findings

Results follow the evaluation protocol in Section 4. The field study first tests whether the pipeline can be applied to measured V2I UWB data. The controlled simulation study then isolates the denoising contribution, downstream effects, anchor-set robustness, and deployment sensitivities under known ground truth.

Field Feasibility on Measured V2I Data

The field track has eight dynamic driving episodes with varying anchor visibility and two static reference logs. The preprocessing pipeline of Section 4 converts asynchronous UWB events into aligned pose–range–mask sequences, so the simulation-trained denoisers and shared downstream solvers apply directly without interface changes. All methods use the same per-anchor affine input calibration. The proposed denoiser is also reported with a degree-2 output calibration on its field outputs. Because this calibration uses field reference information, the corresponding result should be interpreted as a calibrated field diagnostic.

Signal-level range accuracy.

The signal-level columns of Table 4 report range error on the dynamic field episodes, weighted by valid measurement count across all eight episodes. With the shared input calibration and the additional degree-2 output calibration described above, the proposed denoiser obtains the lowest MSE ($1.418\,\mathrm{m}^2$) and MAE ($0.750\,\mathrm{m}$). This corresponds to a 66.9% MSE and 50.8% MAE reduction relative to the calibrated raw input ($4.286\,\mathrm{m}^2$ MSE, $1.524\,\mathrm{m}$ MAE). PoseKalman is the closest baseline at $4.261\,\mathrm{m}^2$ MSE and $1.536\,\mathrm{m}$ MAE, essentially at parity with the calibrated raw input. MLP and PoseMLP yield higher errors than the calibrated raw input, indicating that the simulation-trained feed-forward baselines do not match this field distribution under the calibration used here. Because the proposed denoiser includes an output calibration fitted with field reference information, the field numbers should be read as calibrated deployment diagnostics.

Downstream localization and mapping.

The downstream columns of Table 4 report the corresponding geometry metrics, using the same solver as in simulation. The absolute downstream scale is set by the small site footprint and by the fact that the reference distances come from measured site dimensions plus RTK/INS rather than per-frame surveyed truth. Within this regime, the calibrated output of the proposed denoiser gives the lowest Anchor Error_w (12.23 m) and the highest polygon IoU (0.046), the two cone-layout metrics on which it ranks first among the compared field pipelines.

Table 3. Real-world dynamic-episode results. Signal-level range error is weighted by valid measurement count; downstream geometry uses shared solvers. Lower is better except IoU.

Method	MSE ↓	MAE ↓	Anchor Error_w ↓	IoU ↑	Hausdorff ↓
Raw (calibrated input)	4.286	1.524	13.631	0.000	12.395
Kalman (1D)	12.478	2.533	38.827	0.045	56.029
MLP	13.596	3.101	18.088	0.013	29.589
PoseMLP	26.364	4.638	27.062	0.015	71.999
PoseKalman	4.261	1.536	13.989	0.000	14.372
Proposed method (field-calibrated)	1.418	0.750	12.225	0.046	22.858

Representative progressive work zone geometry reconstruction.

Figure 5 compares the reconstructed anchor layouts as the vehicle accumulates observations across one dynamic episode (10–100% of the episode). The proposed method forms a recognizable trapezoid by 20% and stabilizes by 35%. PoseKalman and MLP require more observations before approaching the correct geometry, while PoseMLP and Kalman are less stable in the early fractions and remain less competitive in the aggregate downstream metrics. This representative episode illustrates the qualitative behavior behind the aggregate anchor-error and polygon-IoU results.

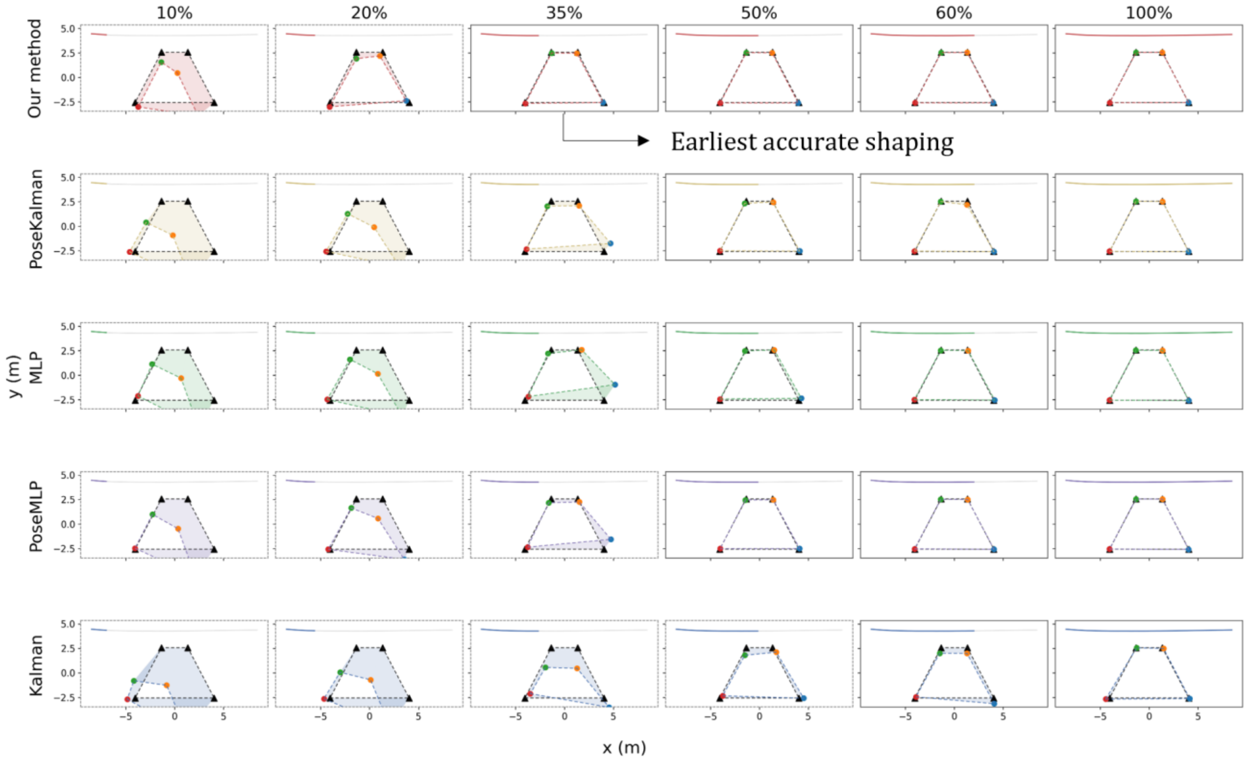

Figure 5. Progressive work zone geometry reconstruction on one real dynamic episode.

Range Denoising in Controlled Simulation

The simulation study uses the challenging and clean regimes of Section 4.2 to isolate denoising performance under controlled conditions. The challenging regime is the target stress case for bursty NLOS and heavy-tailed residuals, whereas the clean regime serves as a negative control showing when simple filtering remains preferable.

Range Denoising Across Challenging and Clean Regimes

Table 4. Range denoising MSE on challenging and clean test splits. Lower is better.

Method	Chall. MSE_LOS ↓	Chall. MSE_NLOS ↓	Chall. MSE_Overall ↓	Clean MSE_LOS ↓	Clean MSE_NLOS ↓	Clean MSE_Overall ↓
Kalman (1D)	1.6736	4.7976	4.1165	0.0360	0.0340	0.0353
MLP	1.3002	3.4651	2.9931	0.2493	0.3207	0.2755
PoseMLP	0.7834	2.6400	2.2352	0.0653	0.1416	0.0933
PoseKalman	0.5636	2.0814	1.7505	0.0313	0.5179	0.2097
Proposed method	0.5285	0.7075	0.6686	0.1042	0.1247	0.1117

The pattern is regime-dependent. In the challenging scenario, the proposed method reduces overall MSE by 77.7% relative to MLP, 83.8% relative to Kalman, and 61.8% relative to PoseKalman, with NLOS MSE 66.0% lower than PoseKalman. In the clean scenario, Kalman remains optimal on overall MSE. But the proposed method still shows competitive performance. This indicates that the proposed architecture is most useful when non-Gaussian NLOS errors and temporal range corruption dominate the measurement process, which appear frequently in the real traffic scenarios.

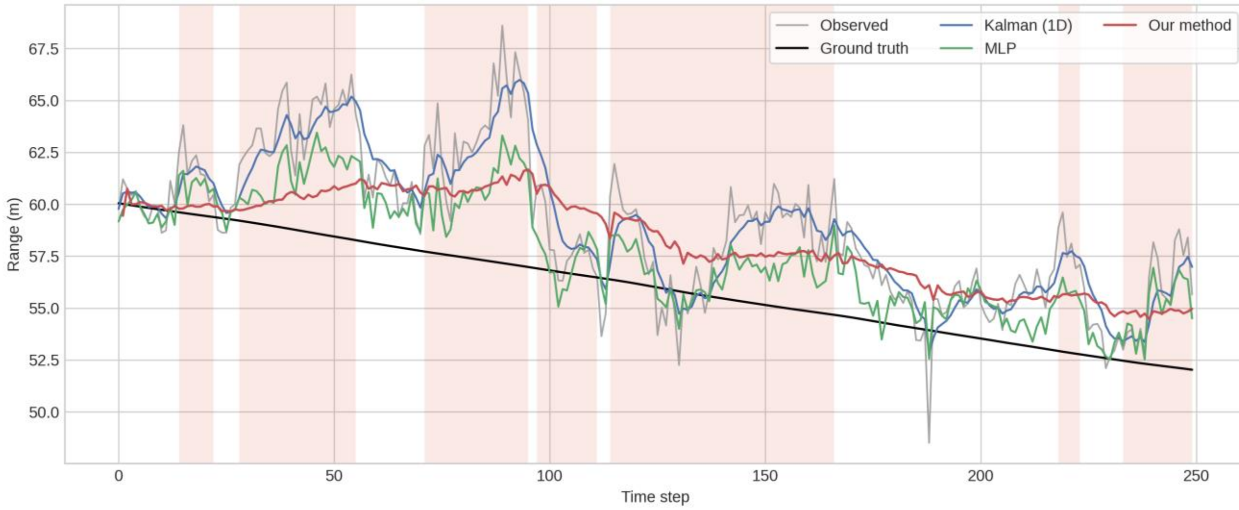

Figure 6. Time-series comparison on one anchor stream from the challenging scenario; shaded regions indicate NLOS-active segments.

Figure 6 gives a sequence-level view of the same effect: across extended NLOS segments, the proposed method suppresses the large positive excursions seen in the raw observations, MLP, and Kalman, and follows the downward ground-truth trend more stably, although residual positive bias remains in some long NLOS intervals.

Polygon-Oriented Model Selection

The pipeline is ultimately judged by downstream geometry rather than by range loss alone, so we select the retained checkpoint on the validation split using polygon IoU under the fixed downstream solver. The challenging-scenario row for the proposed method in Table 5 already uses this polygon-tuned checkpoint. Table 1 compares it with the base checkpoint of the same architecture.

Table 5. Base versus polygon-tuned checkpoints of the proposed method on the challenging scenario test split.

Variant	MSE_Overall ↓	IoU ↑	Hausdorff ↓
Proposed method (base checkpoint)	0.789	0.297	56.65
Proposed method (polygon-tuned checkpoint)	0.669	0.315	53.78

Base vs. polygon-tuned checkpoints of the proposed method on the challenging scenario test split.

The polygon-tuned checkpoint improves the base checkpoint by 15.2% in overall MSE and about 6% in polygon IoU. We therefore use this checkpoint for the remaining challenging-regime analyses. This model-selection rule is used because range MSE is only partially aligned with polygon quality. For the proposed method, the episode-level Spearman correlations of range MSE with polygon IoU and Hausdorff are -0.320 and 0.362, respectively.

Statistical Robustness of Main Results

A single-point MSE can be misleading on a finite test set, so we report episode-level bootstrap confidence intervals (1000 resamples). Table 2 shows the mean overall MSE and the 95% CI for both benchmark regimes. These episode-weighted means need not match the measurement-weighted MSEs of Table 5 exactly, because episodes contain different numbers of valid measurements.

Table 6. Bootstrap robustness for episode-weighted overall MSE, reported as mean episode-level MSE [95% CI].

Method	Challenging scenario ↓	Clean scenario ↓
Kalman	4.173 [4.039, 4.306]	0.0349 [0.0336, 0.0362]
MLP	3.133 [2.957, 3.304]	0.276 [0.260, 0.292]
PoseMLP	2.471 [2.218, 2.745]	0.0916 [0.0845, 0.0989]
PoseKalman	1.760 [1.650, 1.867]	0.191 [0.152, 0.230]
Proposed method	0.665 [0.627, 0.702]	0.111 [0.107, 0.115]

Bootstrap robustness for episode-weighted overall MSE (mean of episode-level MSEs [95% CI]). Best per column in bold.

The challenging-regime confidence interval for the proposed method does not overlap with any baseline, indicating that the main MSE improvement is statistically stable.

Cone Localization, Work Zone Geometry, and Anchor-Set Robustness

This subsection covers cone-position accuracy on the challenging-scenario benchmark and the robustness of cone localization and work zone geometry reconstruction under deployment-relevant anchor perturbations. We first report range-level tail, burst, and stress behavior, then evaluate downstream cone-position and polygon-shape stability under episode-consistent anchor re-indexing and anchor dropout.

Tail, Burst, and Stress Robustness

Table 3 shows that the proposed method has the lightest error tail and the lowest burst max error. Figure 7 plots the corresponding error distribution. Recovery duration is not uniformly minimized, suggesting a trade-off between stable burst suppression and rapid transient correction.

Table 7. Tail percentiles and burst-segment metrics on the challenging scenario test split. Lower is better.

Method	Tail p95 ↓	Tail p99 ↓	Burst max mean ↓	Burst max p95 ↓	Recovery p95 ↓
Kalman	4.140	5.695	3.576	6.847	20
MLP	3.453	5.339	3.075	6.114	10
PoseMLP	2.893	5.095	2.229	5.366	6
PoseKalman	2.480	3.205	1.288	3.185	41
Proposed method	1.644	2.302	1.152	2.574	17

Tail percentiles (overall absolute error, m) and burst-segment metrics on the challenging scenario test split. Lower is better. Best per column in bold.

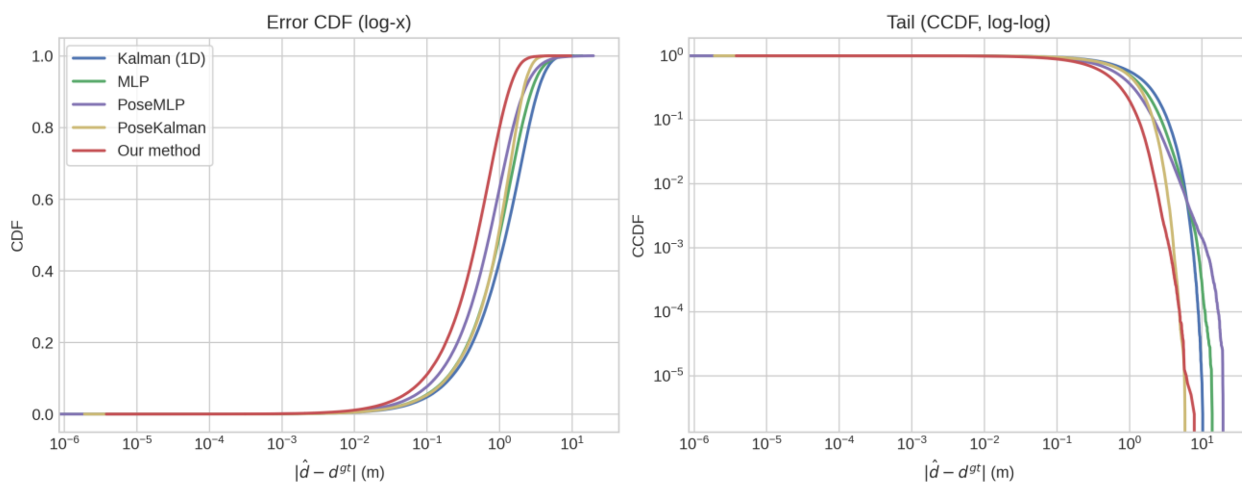

Figure 7. Absolute-error distribution on the challenging scenario, shown as CDF and complementary CDF.

Table 8. Anchor-stress robustness in the challenging scenario: overall MSE under anchor dropout and random episode-consistent permutations.

Method	Drop 0.0 ↓	Drop 0.1 ↓	Drop 0.3 ↓	Drop 0.5 ↓	Perm MSE ↓
Kalman	4.128	4.117	4.143	4.182	4.128 ± 0.000
MLP	3.002	34.942	96.539	193.575	149.909 ± 5.062
PoseMLP	2.242	32.633	127.643	299.089	189.840 ± 6.282
PoseKalman	1.743	1.742	1.723	1.797	1.743 ± 0.000
Proposed method	0.669	0.664	0.687	0.705	0.669 ± 0.000

Anchor-stress robustness in the challenging scenario: overall MSE under anchor dropout rates from 0.0 to 0.5 (mean MSE), and under 20 random episode-consistent permutations (mean ± std across permutations). The Drop 0.0 column is the original stress-test reference and is approximately the same as the nominal challenging-scenario MSE in Table 5, with small differences due to rounding and stress-test evaluation details. The best result in each column is shown in bold.

In Table 4, the Drop 0.0 column applies no perturbation which means the full anchor set is in its canonical order. And therefore it serves as the unperturbed reference corresponding to the nominal challenging-scenario MSE. The remaining columns apply either anchor dropout or anchor re-indexing. The proposed method is insensitive to episode-consistent anchor permutations and remains stable under the anchor-dropout sweep: range MSE stays at 0.669 across all 20 permutations and remains below the baseline values at dropout rates up to 0.5. PoseKalman is also invariant to permutation, and both it and Kalman stay flat under dropout, because all three of these methods process anchors independently or with order-symmetric operators. In contrast, MLP and PoseMLP consume the multi-anchor observation as a fixed-order vector, so episode-consistent re-indexing reshuffles their input and the learned per-slot mapping no longer applies. Their permutation MSE rises to 150–190, and both degrade sharply under dropout as well (193.6 and 299.1 at 50%). PoseMLP is the more brittle of the two despite its lower nominal MSE: conditioning on pose lets it fit the canonical anchor ordering more tightly, which makes it degrade further once that ordering is broken. Figure 8 shows the corresponding dropout curves.

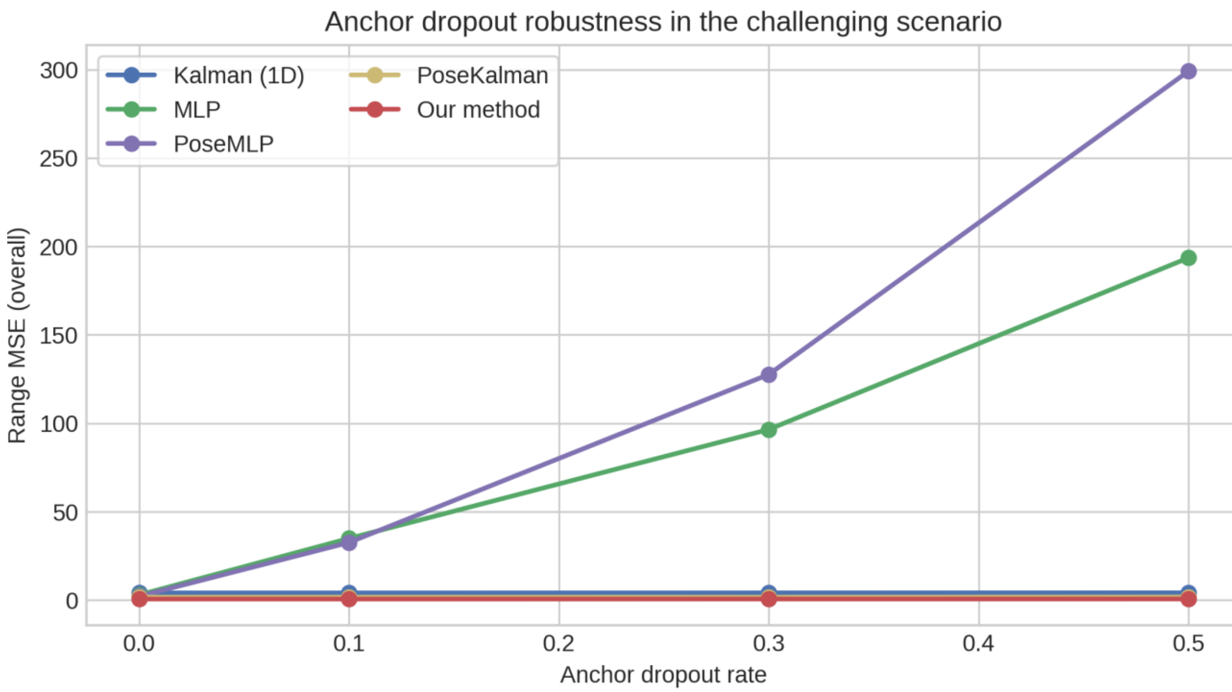

Figure 8. Anchor-dropout robustness on the challenging scenario.

Cone Localization and Work Zone Shape Under Anchor Perturbations

All downstream metrics use the fixed localization pipeline of Section 4.4 with only the input ranges differing across methods. The proposed-method row uses the unified polygon-tuned checkpoint retained above, and baseline rows use their matched retained checkpoints.

Table 9. Cone-position accuracy and work zone shape stability under 20 random episode-consistent anchor permutations on the challenging scenario test split.

Method	Anchor Error_w ↓	IoU ↑	IoU std ↓
Kalman	33.16	0.264	0.000
MLP	32.07	0.221	0.007
PoseMLP	27.83	0.218	0.006
PoseKalman	35.05	0.268	0.000
Proposed method	29.25	0.313	0.001

Cone-position accuracy (Anchor Error_w, under canonical anchor ordering) and work zone shape stability (polygon IoU, mean ± std over 20 random episode-consistent anchor permutations) on the challenging scenario test split. Lower is better except IoU. Best per column in bold, second best underlined.

Table 10. Downstream robustness under anchor dropout on the challenging scenario. Polygon IoU is averaged over non-degenerate episodes.

Method	IoU@0.0 ↑	IoU@0.3 ↑	IoU@0.5 ↑
MLP	0.284	0.116	0.052
PoseKalman	0.268	0.213	0.172
Proposed method	0.315	0.236	0.230

Downstream robustness under anchor dropout on the challenging scenario. Polygon IoU is reported for selected dropout rates. Best per column in bold. Polygon IoU is averaged only over non-degenerate episodes with at least three retained mapped anchors.

For compactness, Table 6 reports a representative subset: the index-sensitive MLP, the strongest invariant baseline PoseKalman, and the proposed method.

Table 5 reports cone-position accuracy under the canonical anchor ordering (Anchor Error_w) alongside work zone shape stability under permutation (polygon IoU mean ± std over 20 random re-indexings). PoseMLP has the lowest canonical Anchor Error_w at 27.83 m, with the proposed method second at 29.25 m, but this canonical-order advantage is not robust to re-indexing: PoseMLP's polygon IoU under permutation drops to 0.218 with non-negligible std (0.006), whereas the proposed method preserves polygon IoU at 0.313 with std 0.001 and ranks first on the work zone shape metric. PoseKalman is similarly invariant. Anchor dropout gives a more nuanced picture (Table 6): at 30% dropout the proposed method has the highest polygon IoU (0.236) among the reported methods, but at 50% dropout the polygon quality degrades for every method even though range MSE for the equivariant denoisers stays flat. This is observability-limited. 28.2% of episodes retain only two anchors after 50% masking and become weakly constrained for cone localization (recovering an anchor position from range measurements, which needs several well-spread tag observations). And even here the proposed method retains the highest polygon IoU (0.230 vs 0.172 for PoseKalman and 0.052 for MLP).

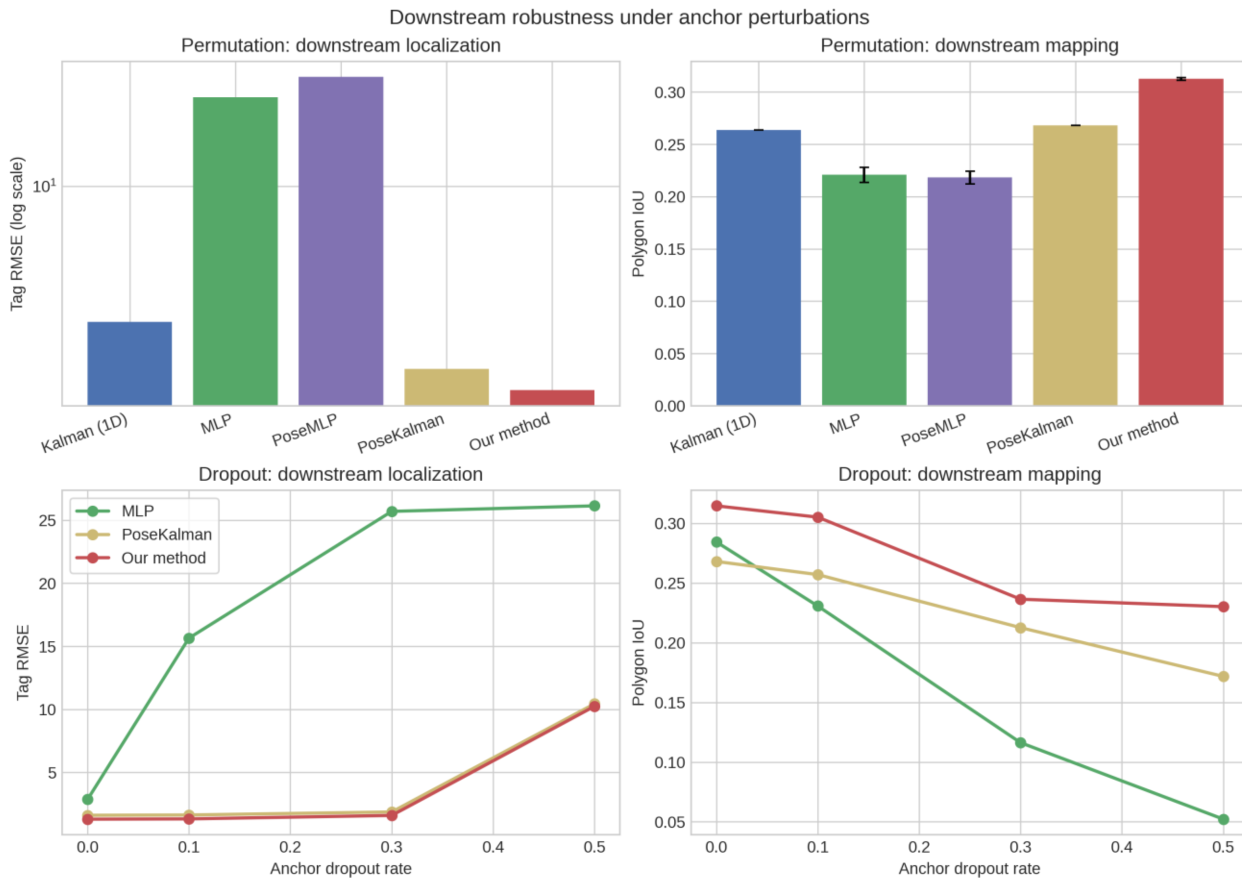

Figure 9. Downstream robustness under anchor perturbations on the challenging scenario.

Figure 9 visualizes both regimes side by side.

Component and Deployment Sensitivity

The final analyses examine how strongly the retained model depends on its architectural information pathways and on the quality of the pose input. These tests should be interpreted as component and deployment sensitivities rather than as additional primary benchmarks.

Component Ablation

We measure the contribution of each design choice by ablating it on the unified challenging-scenario checkpoint: pretraining removed before fine-tuning, set context removed at runtime, and pose input removed at runtime. The context and pose rows are runtime-removal sensitivities, so they indicate how strongly the trained model relies on each information pathway rather than serving as fully retrained architectural ablations. Table 7 reports the resulting metrics.

Table 11. Component ablation on the challenging scenario test split.

Variant	MSE_Overall ↓	MSE_NLOS ↓	IoU ↑
Full model	0.669	0.707	0.315
w/o pretraining	0.937	1.008	0.286
w/o context (runtime)	1.079	1.064	0.244
w/o pose (runtime)	6.810	7.741	0.131

Component ablation on the challenging scenario test split (family of the proposed method). Best per column in bold.

Pose conditioning is the largest observed dependency in this runtime-removal sensitivity test: removing it at runtime raises overall MSE by an order of magnitude (from 0.669 to 6.810) and substantially degrades the trained model relative to the other tested variants. Set context is the next largest observed dependency, followed by predictive pretraining. This ordering holds for both overall MSE and polygon IoU. The "w/o pretraining" row also quantifies the design role of Stage 1 described in Section 3.1: omitting Stage 1 and training the same backbone from random initialization under supervised loss raises overall MSE by 40.1% and NLOS MSE by 42.6%, with the largest benefit in the NLOS-heavy regime where the supervised loss alone is most non-convex.

Pose-Noise Sensitivity

Table 12. Pose-noise sensitivity on the challenging scenario when zero-mean planar noise is injected into the model pose input at test time.

Method	MSE@0.0 ↓	MSE@0.25 ↓	MSE@0.5 ↓
PoseMLP	2.242	2.250	2.274
PoseKalman	1.743	2.739	3.415
Proposed method	0.669	0.727	0.912

Pose-noise sensitivity on the challenging scenario when zero-mean planar noise is injected only into the model pose input at test time. Best per column in bold.

Table 8 isolates the dependence on pose quality. It lists only the pose-conditioned methods, since pose-free Kalman and MLP are unaffected by perturbations applied to the pose input. For range MSE, the proposed method stays the most accurate across the injected-noise levels in the table: overall MSE rises from 0.669 to 0.727 at 0.25 m and to 0.912 at 0.50 m. PoseMLP is flatter but markedly less accurate, while PoseKalman is more sensitive to corrupted pose. Polygon IoU is only partially aligned with range denoising: the proposed method's IoU decreases with pose noise, from 0.315 to 0.298 at 0.50 m. The larger 1.0 m perturbation point is shown in Figure 10 and omitted from the compact table. Moderate pose error therefore does not invert the range MSE ranking among these methods. Figure 10 shows the curves.

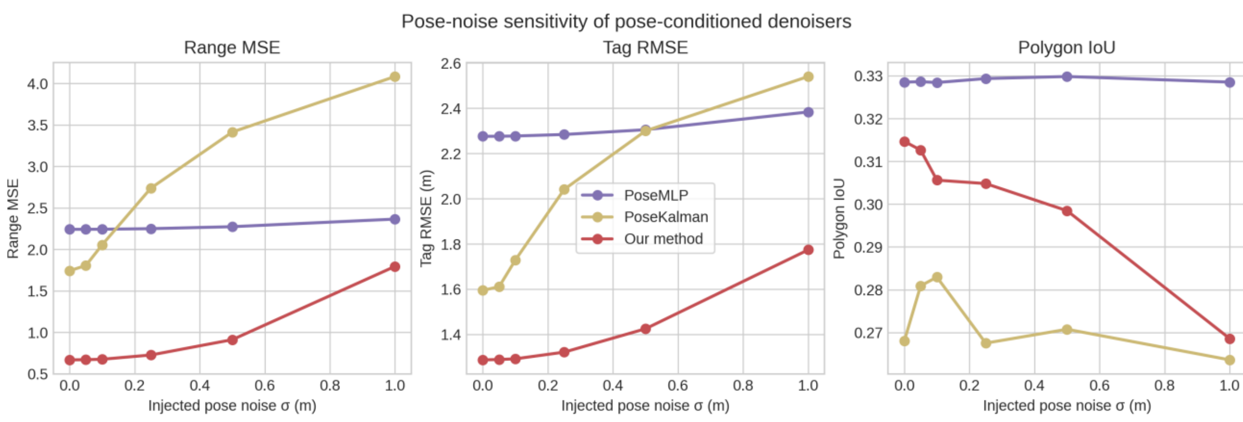

Figure 10. Pose-noise sensitivity of pose-conditioned denoisers on the challenging scenario.

7. CRM System Prototype and Baseline Localization Evaluation

This chapter documents the low-cost UWB-based CRM system assembly, two-anchor and four-anchor localization logic, and the static and dynamic experiments that establish the baseline feasibility of CRM-enabled lane-level positioning for cooperative automated driving.

7.1 System Architecture and UWB Ranging

The baseline CRM system uses ESP32 UWB modules as connected reference markers. In a roadside deployment, fixed CRM anchors are placed at known positions and a tag CRM is mounted on the vehicle or mobile unit. A local computer collects the UWB ranging data, manages communications, and executes the positioning calculations. This architecture corresponds directly to the proposal task on system assembly because it demonstrates the roadside-device, onboard-unit, and software components needed for a low-cost CDA localization layer.

Distance measurement relies on UWB time-of-flight ranging. In single-sided two-way ranging, one device sends a signal and receives a delayed response, while double-sided two-way ranging introduces an additional exchange to reduce clock-related timing error. The experiments use these UWB measurements as the basic distance observations for subsequent triangulation.

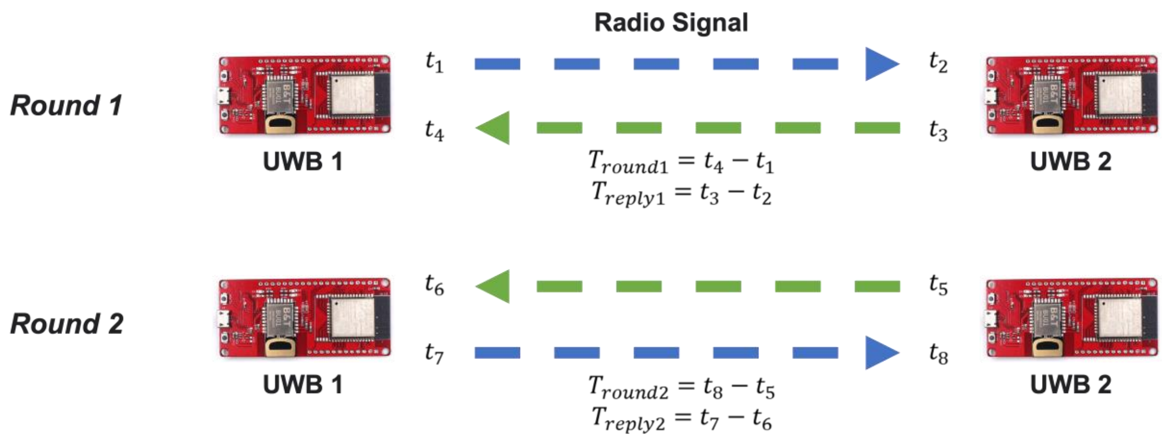

Figure 11. UWB two-way ranging process used by the CRM distance-measurement module.

7.2 Triangular Positioning with Two and Four Anchors

The prototype paper evaluates two positioning configurations. The basic triangular method uses two fixed anchors and one moving tag; the tag position is estimated from the measured distances to the two anchors and the known anchor baseline. The advanced method uses four anchors placed in a rectangular layout. It forms multiple anchor-pair triangles, computes several candidate tag positions, and averages or smooths the estimates to improve robustness against individual measurement errors.

The four-anchor method is most relevant to work zone deployment because cone or marker layouts can be arranged around the work area and because multiple anchor pairs provide redundancy when one pair is affected by noise, obstruction, or poor geometry.

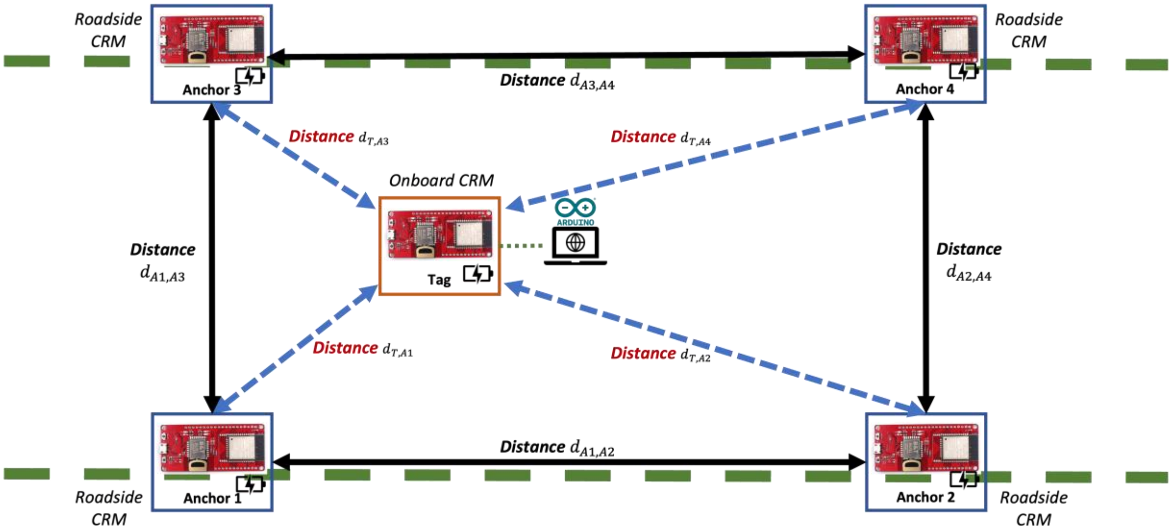

Figure 12. Four-anchor CRM/Tag architecture for advanced triangular positioning.

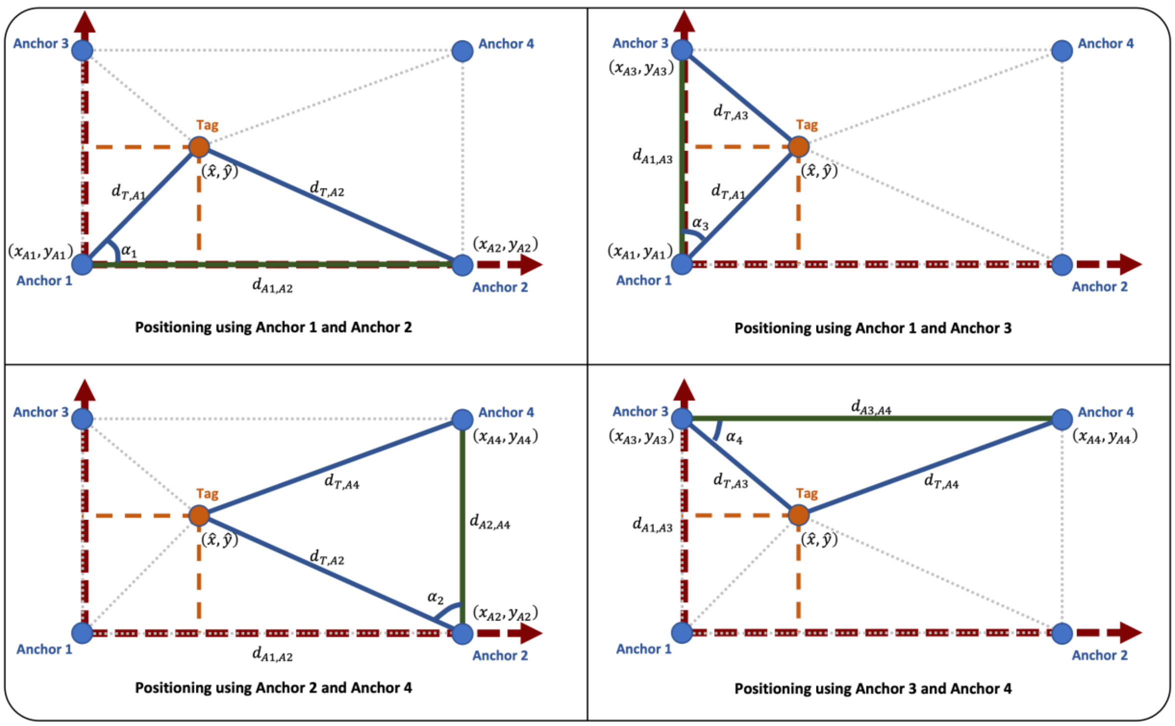

Figure 13. Four-anchor triangular positioning geometry and coordinate averaging.

7.3 Experiment Design

The prototype experiment used four CRM units as anchors and one CRM unit as the tag. The test environment was an outdoor rectangular area measuring approximately 13.15 m by 7.5 m. Static tests placed the tag at fixed locations for 60-second sessions, while dynamic tests moved the tag horizontally along a straight line near Y = 3.75 m at roughly 0.48 m/s for 30-second sessions. These experiments were designed to evaluate both distance measurement precision and two-dimensional positioning performance under controlled but realistic field conditions.

7.4 Static Scenario Findings

The static scenario showed that CRM distance and coordinate errors were concentrated near zero and generally remained below 0.1 m for both X and Y coordinates. The right-skewed error distributions indicate that most measurements are accurate, with occasional over-estimation likely caused by reflections, signal-strength variation, or local environmental effects. Two-dimensional coordinate estimates clustered close to the ground-truth positions, supporting the feasibility of the CRM system for precise static localization.

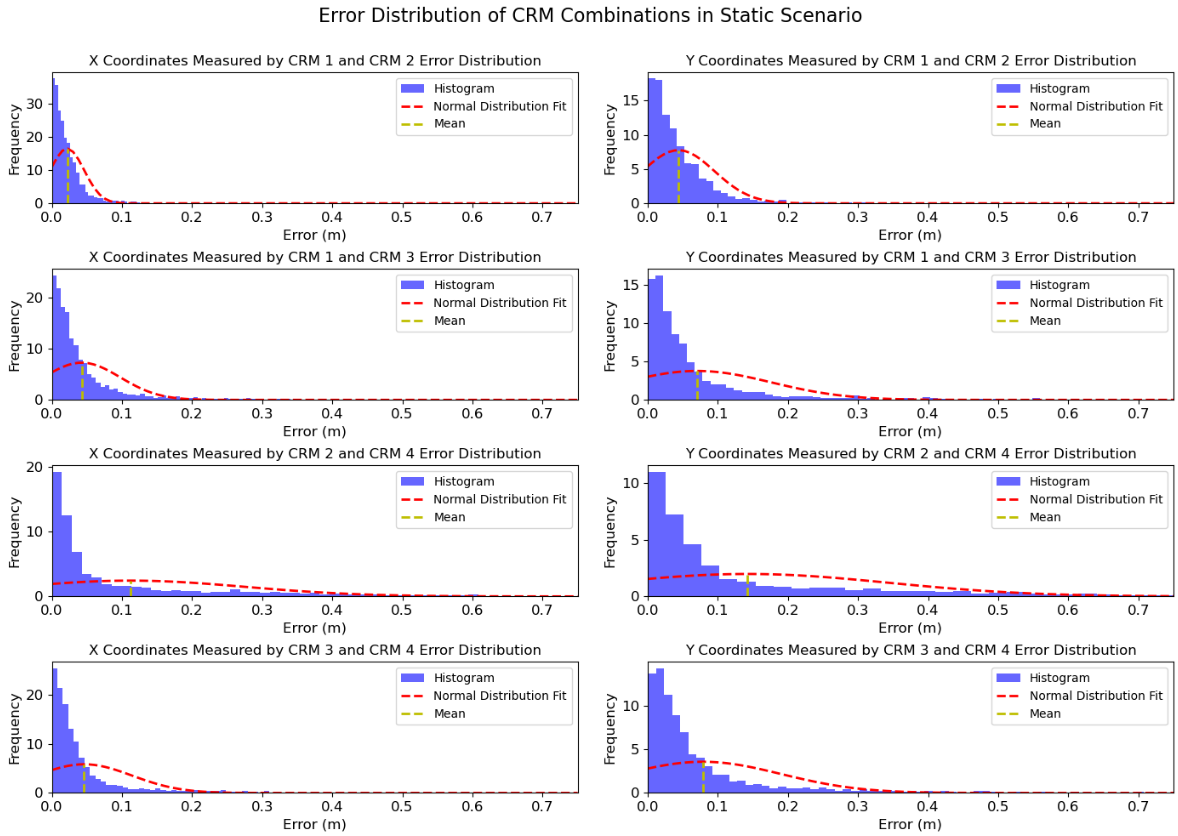

Figure 14. Error distribution of CRM combinations in the static scenario.

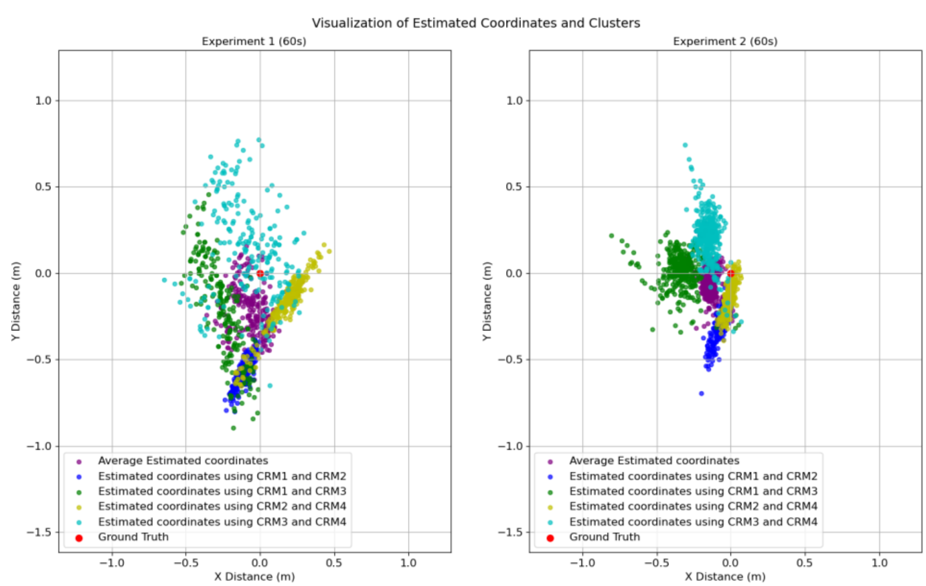

Figure 15. Static-scenario two-dimensional positioning clusters and ground-truth comparison.

7.5 Dynamic Scenario Findings

The dynamic scenario maintained mean errors typically below 0.2 m, indicating that the low-cost CRM/UWB setup can track moving tags with useful accuracy. The experiments also revealed more pronounced Y-axis fluctuations than X-axis fluctuations, particularly under rapid motion or when anchor geometry and signal propagation were less favorable. The smoothed moving traces nevertheless followed the ground truth closely, showing the value of aggregation and temporal smoothing for real-time deployment.

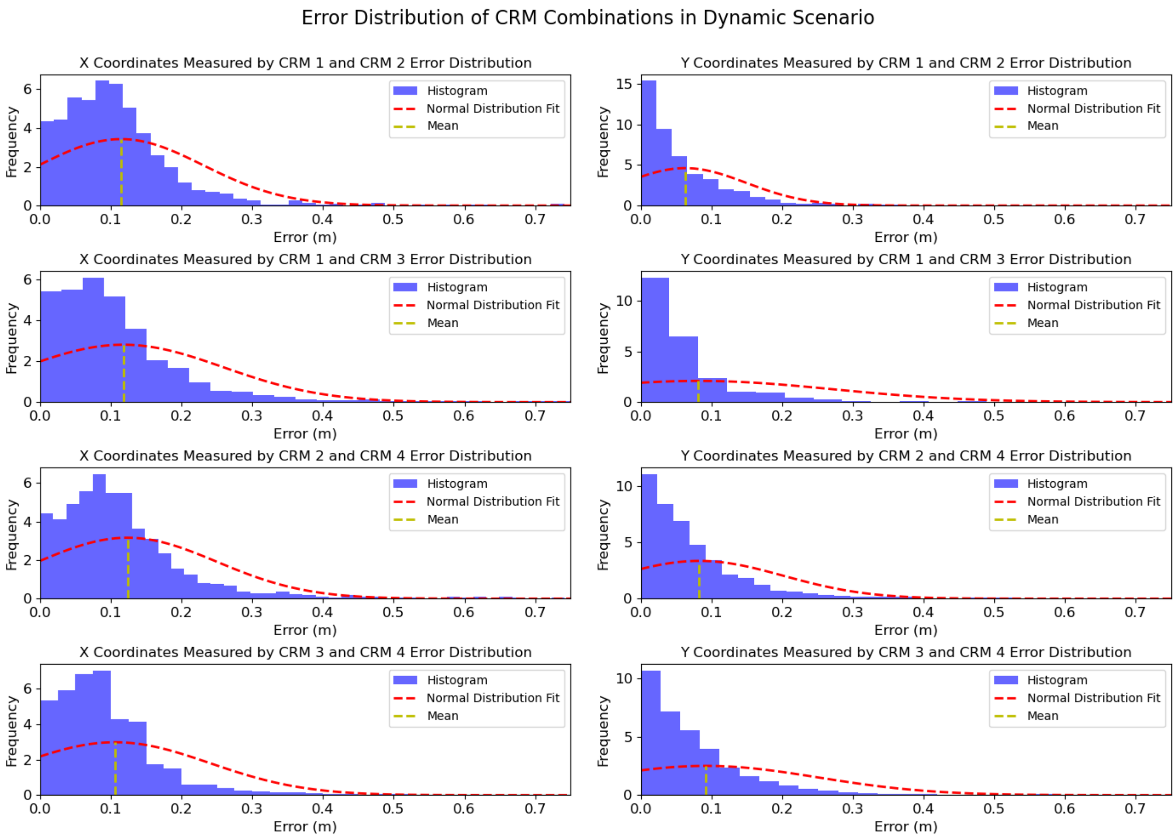

Figure 16. Error distribution of CRM combinations in the dynamic scenario.

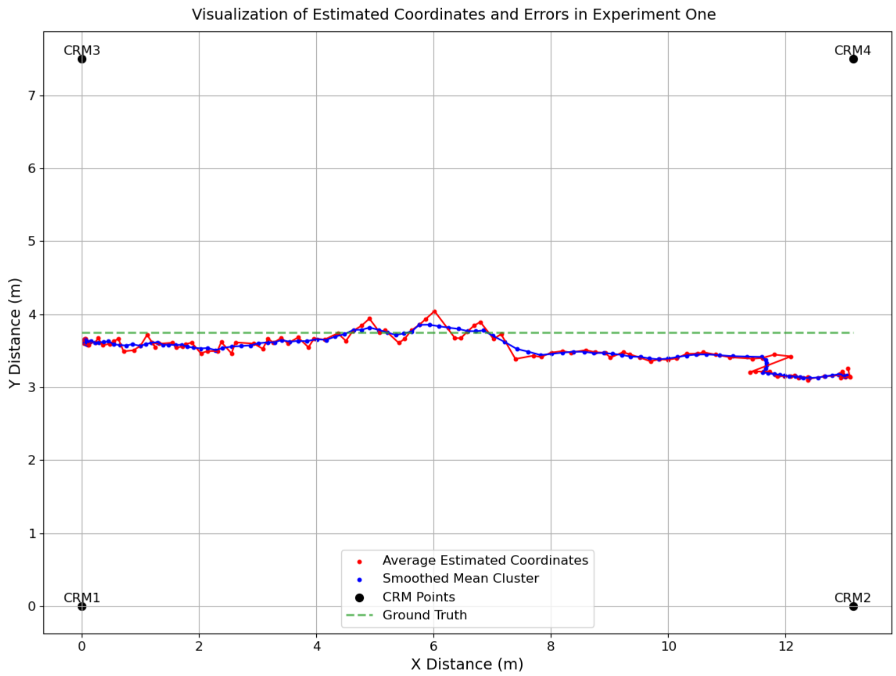

Figure 17. Estimated moving trace in Dynamic Experiment 1.

Figure 18. Estimated moving trace in Dynamic Experiment 2.

7.6 Deployment Implications

The prototype study provides the baseline evidence required by the proposal tasks on CRM distance measurement calibration and vehicle localization performance evaluation. It demonstrates that the CRM concept can produce low-cost, lane-level positioning information, while identifying calibration needs for dynamic motion, anchor placement, and Y-axis sensitivity. These findings motivate the later denoising and work zone mapping work because robust operation in traffic work zones requires both accurate raw ranging and downstream correction methods for NLOS, multipath, and occlusion.

8. CRM-Enhanced WZDx and LLM-Based Work Zone Navigation Refinement

This chapter addresses Proposal Task 5 by treating the CRM system as a high-resolution data source for WZDx-style work zone data provision. The central idea is to combine WZDx semantic information with CRM-derived cone geometry so that automated vehicles can receive both lane-closure intent and physically accurate corridor geometry.

8.1 Motivation and Input-Output Formulation

WZDx feeds provide standardized work zone descriptions, lane status metadata, and rough polyline geometries. These data are valuable for planning and traveler information, but the official polyline can be offset from the actual lane geometry or lack the cone-level detail needed by AV planning stacks. The CRM system addresses this gap by localizing cone or barrel markers with UWB ranging and providing high-resolution geometric anchors for the active work zone.

The WZDx/LLM framework maps five inputs to three outputs. The inputs are travel direction, WZDx natural-language description, structured lane list, rough WZDx polyline, and CRM-derived cone/barrel coordinates. The outputs are human-readable navigation instructions, a list of recommended open lanes, and a refined GeoJSON LineString representing the centerline of the safe drivable corridor.

8.2 CRM Cone Localization as a WZDx Data Source

The framework assumes that cone or barrel positions are obtained through the CRM localization workflow: a mobile tag with known trajectory measures UWB distances to fixed cone-mounted anchors, and a least-squares localization process recovers the cone coordinates. These coordinates then refine WZDx by improving lane-closure geometry, start/end alignment, and the physical boundary definition of the work zone.

8.3 LLM Prompt Template and Reasoning Guardrails

A structured prompt is used to make the reasoning task explicit. The prompt instructs the model to identify open and closed lanes, treat the supplied WZDx polyline as a rough sketch rather than ground truth, use CRM cone coordinates as the physical boundary, and generate a smooth center-aligned polyline that follows the direction of travel while remaining within open general-purpose lanes. This design makes the output interpretable and auditable: a human operator can check the recommended lanes, the narrative instruction, and the generated coordinates against the original WZDx and CRM inputs.

8.4 Case Study 1: Right-Lane Closure on State Highway 113

The first case study considers a standard northbound right-lane closure on State Highway 113. The WZDx polyline is substantially offset from the actual roadway and cone line. The CRM coordinates provide the physical cone boundary, allowing the LLM to identify the right general-purpose lane as closed, recommend the remaining open lane, and produce a refined green centerline aligned with the real work zone geometry.

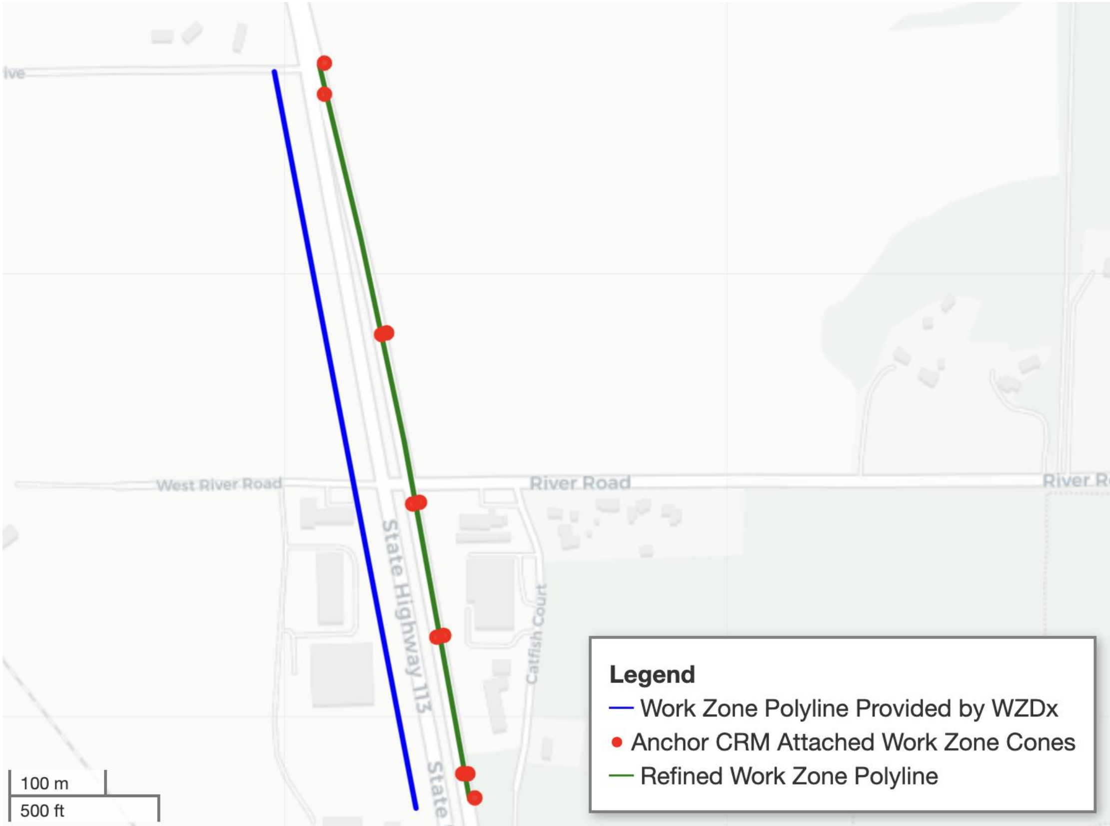

Figure 19. CRM-enhanced refinement of a WZDx right-lane closure: overview map.

Figure 20. CRM-enhanced refinement of a WZDx right-lane closure: zoomed map.

8.5 Case Study 2: Curved Multi-Lane Closure on I-41 Northbound

The second case study considers a more complex curved work zone with three adjacent left general-purpose lanes and the left shoulder closed. The original WZDx polyline drifts from the roadway geometry, while the CRM cone coordinates trace the true work zone boundary. The LLM correctly identifies the rightmost general-purpose lane as the usable lane and generates a smoothed refined polyline that follows the curvature of the open corridor.

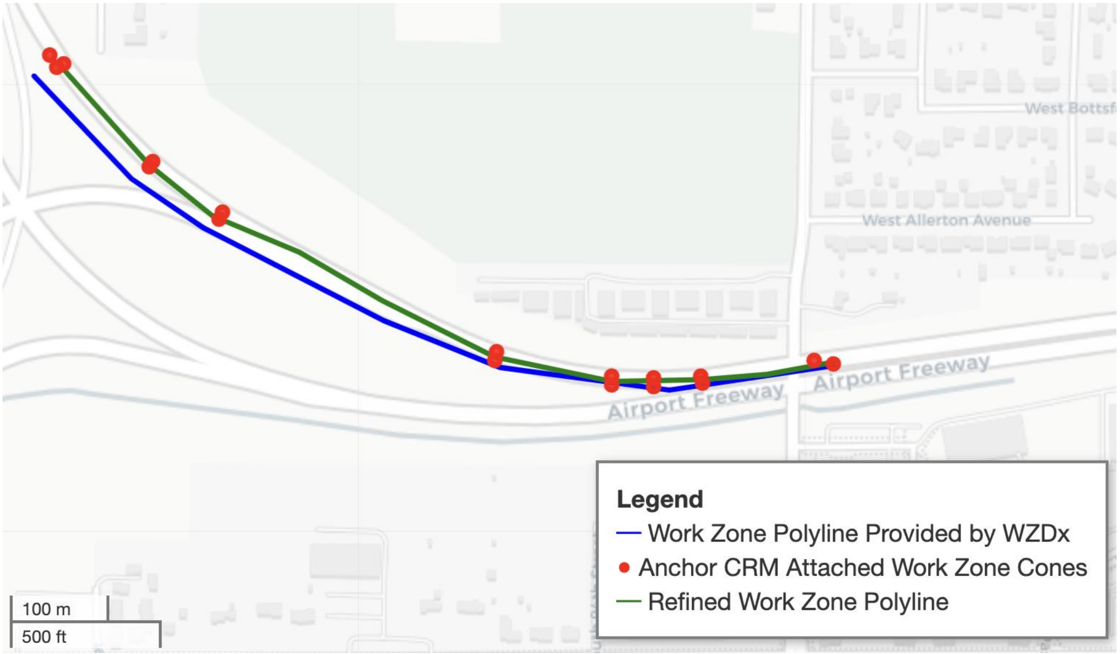

Figure 21. CRM-enhanced refinement of a curved multi-lane closure: overview map.

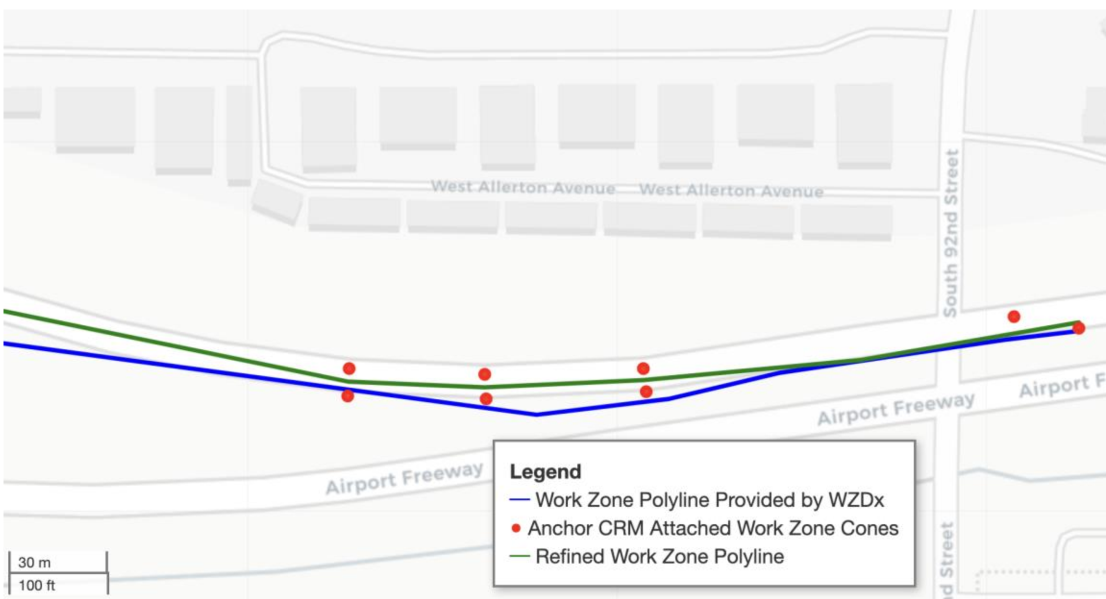

Figure 22. CRM-enhanced refinement of a curved multi-lane closure: zoomed map.

8.6 Findings and Technology-Transfer Value

The WZDx/LLM study demonstrates a practical pathway from CRM sensing to work zone data services. CRM measurements provide the missing geometric fidelity; WZDx provides the policy and semantic structure; and the LLM prompt converts both into an AV-ready representation. The case studies show that the method can correct polyline misalignment and produce interpretable lane recommendations. For deployment, the LLM output should be treated as a candidate product that requires schema validation, geometric consistency checks, and human or automated quality assurance before being broadcast to vehicles or submitted as a production work zone data feed.

9. Integrated Technical Conclusions

The integrated project supports a coherent CRM deployment pathway: first, a low-cost UWB hardware prototype can produce lane-level distance and localization measurements; second, advanced denoising and geometry reconstruction can convert noisy multi-anchor field ranges into robust cone and work zone polygon estimates; and third, CRM-derived cone geometry can enrich WZDx feeds and support AV-ready navigation instructions and refined work zone centerlines.

The results also identify the remaining engineering requirements for acceptance and deployment. The prototype study needs broader field calibration across speeds, anchor layouts, and environmental conditions. The denoising study shows strong benefits in NLOS and heavy-tailed regimes but requires larger field datasets for production calibration. The WZDx/LLM study demonstrates a promising data-feed pathway but requires deterministic validation, schema checks, and human-in-the-loop review before operational use.

Overall, the final report satisfies the proposal's emphasis on a low-cost roadside device system that supports CAV localization and detailed work zone geometry provision. The work advances the CRM system from prototype hardware toward a multi-layer CDA data source: distance measurement, localization, boundary reconstruction, and WZDx-compatible work zone information provision.

10. Recommendations

- Maintain the CRM/UWB system as the primary low-cost gateway architecture for infrastructure-supported CDA. Future deployments should prioritize four-anchor or redundant-anchor layouts to improve observability and robustness.
- Standardize field calibration protocols for static and dynamic CRM localization. The prototype results show low errors in controlled conditions but also reveal Y-axis and motion-related sensitivity that should be addressed before large-scale deployment.
- Use the pose-conditioned, permutation-equivariant denoising backbone in work zone deployments where NLOS, multipath, unordered anchor returns, and partial anchor visibility are expected. Retain Kalman and other conventional filters as interpretable diagnostic baselines.
- Define minimum observability requirements for work zone mapping, including minimum visible anchors, anchor spacing, tag trajectory coverage, and fallback behavior when polygon reconstruction becomes degenerate.
- Translate CRM cone and polygon outputs into WZDx/CWZ-compatible entities, including work zone event identifiers, geometry, timestamps, source type, confidence, and update frequency. This should be the next technology-transfer step for Task 5.
- Treat LLM-generated WZDx refinement as a candidate data product rather than an unverified production feed. Require schema validation, geometric consistency checks, lane-status consistency checks, and human-in-the-loop review for safety-critical use.
- Archive all raw data, processed sequences, code/notebooks, prompt templates, case-study inputs, generated outputs, and QA checklists in a dataset package with a manifest and checksums so CCAT can verify project outputs and reproduce the reported findings.

11. Final Documentation of Outputs, Outcomes, and Impacts

This section is structured to satisfy the final reporting instruction categories for outputs, outcomes, and impacts, while matching the 2024 proposal statement that the project should deliver a CRM system, experimental datasets, and a project report with system configurations, methodologies, results, and actionable recommendations.

Table 13. Output documentation by required reporting category.

Required output category	Project documentation	Status
Presentations	"Roadside Device System for CDA", Aug 18 2025, Yang Cheng, Steven T. Parker, Bin Ran, Rui Gan, Jiaxi Liu, Junwei You, 2025 CCAT Global Symposium	Completed

Required output category	Project documentation	Status
	"Prototyping a Low-cost Roadside Device System for Cooperative Automated Driving", Oct 24 2025, Yang Cheng, Steven T. Parker, Bin Ran, Rui Gan, Jiaxi Liu, Junwei You, CCAT Safety Working Group Meeting "A SELF-SUPERVISED APPROACH FOR ROADWAY WORK ZONE LOCALIZATION", Jan 2026, Jiaxi Liu, Yang Cheng, Rui Gan, Junwei You, Weizhe Tang, Bin Ran, TRB 2026	
Journal papers / reports / publications	"Vehicle-to-Infrastructure Work Zone Geometry Reconstruction with Pose-Conditioned UWB Range Denoising.", Jiaxi Liu, Yang Cheng, Rui Gan, Junwei You, Weizhe Tang, Bin Ran "A Unified LLM Framework for Work Zone Data Provision: Enhancement for Work Zone Data Exchange Feeds", Rui Gan, Junwei You, Jiaxi Liu, Weizhe Tang, Bin Ran	Submitted
Policy papers	No standalone policy paper is claimed. The report provides deployment guidance relevant to low-cost roadside CDA, WZDx/CWZ data-source discussions, work zone digital infrastructure, and agency data-quality review.	N/A
New courses / educational material	No new course is claimed. The hardware, denoising, mapping, WZDx, and LLM prompt examples can be converted into graduate modules on CAV localization, V2I sensing, work zone data standards, and AI-assisted transportation data provision.	N/A
Websites or internet sites	No public project website is reported. A dataset landing page or repository should be created if CCAT requires public dissemination.	N/A
New methodologies, technologies, or techniques	Low-cost CRM/UWB prototype; static and dynamic UWB triangulation workflow; anomaly detection and smoothing for baseline localization; pose-conditioned permutation-equivariant UWB denoising; downstream cone/polygon reconstruction; CRM-enhanced WZDx/LLM prompt template for refined work zone navigation.	Completed research output.
Inventions, patents, and licenses	The proposal identifies CRM as an existing technology/IP basis. No new patent or license filing is claimed in this report unless separately confirmed by the project team.	N/A
Data or databases	Raw data files have been submitted to Dryad with a DOI assigned (https://doi.org/10.5061/dryad.xgxd254xf).	Completed
Software / analytical models	UWB ranging and triangulation scripts; static/dynamic smoothing and anomaly detection; denoising model and baselines; cone localization and polygon reconstruction solvers; WZDx/LLM prompt and output-validation scripts/notebooks.	Completed
Equipment / instruments / research material	Cone-mounted UWB/CRM units, vehicle/mobile tag CRM, power supply setup, local computer, RTK/INS reference system for V2I study, work zone proxy layouts, and experiment protocols.	Completed
Students / trainees	Graduate and student researchers contributed to hardware/software setup, field experiments, data processing, model development, WZDx/LLM case studies, analysis, and report preparation.	All project team members are listed as authors in this report

Table 14. Outcome and impact documentation.

Outcome / impact category	Project evidence	Interpretation
Increased understanding and awareness	The integrated report clarifies how low-cost CRM devices can support CAV localization, work zone boundary mapping, and WZDx-ready work zone data provision.	Improves understanding of infrastructure-supported CDA for difficult ODDs.
Increase in body of knowledge	The project contributes baseline CRM localization evidence, a formal V2I geometry-inference/denoising framework, and a semantic-geometric WZDx/LLM refinement workflow.	Adds hardware, algorithmic, and data-standard knowledge.
Improved processes, technologies, techniques, and skills	The report documents a repeatable pipeline from UWB ranging to localization, denoising, cone mapping, polygon reconstruction, and WZDx-compatible output generation.	Provides a technical workflow for future CRM/CDA evaluation.
Enlargement of trained transportation professionals	Student researchers gained experience in UWB hardware, field testing, simulation, machine learning, work zone geometry, WZDx data elements, LLM prompt design, and technical reporting.	Supports workforce development.
Adoption of new technologies, techniques, or practices	The work identifies a path for agencies and contractors to use low-cost CRM markers as a data source for CAV localization and work zone feeds. Full operational adoption is not claimed.	Potential adoption pathway identified.
Impacts on operation and safety of transportation system	More accurate and timely work zone geometry can improve CAV situational awareness, provide redundancy beyond onboard perception, and reduce navigation uncertainty in constrained work zones.	Potential safety and mobility impact; deployment study needed for quantified benefit.
Mobility and work zone management impacts	CRM-enhanced WZDx outputs can support clearer open-lane guidance, better lane-closure alignment, and more reliable AV/crowdsourced work zone navigation.	Potential mobility and work zone operations benefit.
Capital / operating cost impacts	Low-cost UWB/CRM devices may reduce reliance on survey-grade	Potential cost-reduction

Outcome / impact category	Project evidence	Interpretation
	mapping or expensive smart work zone devices for temporary geometry updates.	impact; needs cost analysis.
Physical, institutional, and information resources	Dataset schemas, prompt templates, evaluation code, and data dictionaries strengthen information resources for future CCAT, WisDOT, and research use.	All project data have been archived for future use and reproducibility

12. Experimental Dataset Documentation and Readiness Package

The following dataset covers all three technical outputs and is available in the Dryad repository: https://doi.org/10.5061/dryad.xgxd254xf. This dataset includes the actual raw CRM logs, V2I UWB logs, processed sequences, WZDx case-study files, LLM prompt/output records, scripts, and checksums.

Table 15. Experimental dataset acceptance-ready package contents.

Dataset component	Description	Acceptance status
README.md	Dataset purpose, project citation, contact, license/access restrictions, folder structure, preprocessing summary, reproduction instructions, and relationship among the three technical studies.	Uploaded to Dryad
dataset_manifest.csv	One row per file with filename, folder, data type, description, source study, collection source, date/time, version, checksum, and required/not-required status.	Uploaded to Dryad
data_dictionary.csv	Field-level definitions for CRM static/dynamic logs, UWB events, pose streams, masks, anchors, calibration, denoised outputs, downstream metrics, WZDx inputs, prompt templates, and LLM outputs.	Uploaded to Dryad
crm_prototype/raw/	Raw four-anchor CRM/UWB static and dynamic experiment logs, tag/anchor configuration files, and experiment metadata.	Uploaded to Dryad
crm_prototype/processed/	Triangulated coordinates, smoothed traces, error distributions, anomaly flags, and baseline localization metrics.	Uploaded to Dryad
v2i_uwb/raw/	Raw UWB event logs, RTK/INS pose streams, static reference logs, and dynamic work zone geometry episodes.	Uploaded to Dryad
v2i_uwb/processed/	Aligned 0.10 s pose-range-mask sequences, geometry-derived reference ranges, calibration parameters, train/validation/test split identifiers, denoised ranges, and downstream geometry outputs.	Uploaded to Dryad
wzdx_llm_cases/	WZDx event descriptions, structured lane lists, rough WZDx polylines, CRM cone/barrel coordinates, prompt templates, LLM responses, recommended lanes, refined polylines, and validation notes.	Uploaded to Dryad
results/	CSV versions of reported range, localization, robustness, downstream, ablation, pose-noise, CRM prototype, and WZDx/LLM case-study outputs.	Uploaded to Dryad
qa_checklist.csv	Checklist for missing values, time alignment, units, coordinate frames, anchor IDs, split integrity, WZDx schema consistency, geometry validity, LLM output validation, and reproducibility.	Uploaded to Dryad
code_or_notebooks/	Evaluation scripts or notebooks for reproducing the figures, tables, prompt outputs, and data products.	Uploaded to Dryad

13. Challenges and Lessons Learned

Table 16. Challenges and lessons learned.

Challenge	Observed issue	Lesson learned / response
Prototype field scale	The CRM prototype tests used a compact outdoor rectangular area and controlled static/dynamic trajectories.	Use results as feasibility evidence and expand to larger work zone geometries before operational claims.
Dynamic localization sensitivity	Dynamic tests showed stronger Y-axis fluctuations and occasional deviations from the ground truth.	Improve calibration, anchor placement, smoothing, and real-time processing before deployment.
NLOS and bursty UWB errors	NLOS bias and burst outliers dominate challenging work zone UWB regimes and degrade raw range usability.	Use pose-conditioned temporal denoising and robust supervision in realistic deployments.
Anchor order and missing anchors	Raw ranging interfaces can return unordered anchor lists, and anchors can be occluded or unavailable.	Use permutation-equivariant and mask-aware model designs as deployment requirements.
Severe anchor dropout	At high dropout rates, multilateration and polygon reconstruction become observability-limited.	Set minimum visible-anchor and geometry-spread requirements for deployment.
Field calibration dependency	The best field denoising diagnostic uses degree-2 output calibration with reference distances.	Collect larger field datasets to learn robust calibration without overfitting a small site.
WZDx geometric ambiguity	Rough WZDx polylines can be offset from actual cone/lane geometry.	Use CRM cone coordinates as high-resolution physical anchors for data-feed refinement.

Challenge	Observed issue	Lesson learned / response
LLM validation and safety assurance	LLM outputs can be useful but require deterministic checks before safety-critical use.	Validate lane status, geometry, schema, direction, and clearance; use human-in-the-loop review.
Dataset readiness	Raw data were not included for all reported studies in the uploaded materials.	The report provides a manifest and schema; raw logs, prompts, code, checksums, and outputs should be attached before submission.

14. Acknowledgments

This study is supported by the Center for Connected and Automated Transportation (CCAT), University of Michigan, through Grant No. 69A3552348305. The views expressed are those of the authors.

15. References

Alarifi, Abdulrahman, Abdulmalik Al-Salman, Mansour Alsaleh, Abdullah Alnafessah, Sami Al-Hadhrami, M. A. Al-Ammar, and Hend S. Al-Khalifa. 2016. "Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances." Sensors 16 (5): 707. .

Angarano, Simone, Vittorio Mazzia, Francesco Salvetti, Giovanni Fantin, and Marcello Chiaberge. 2021. "Robust Ultra-Wideband Range Error Mitigation with Deep Learning at the Edge." Engineering Applications of Artificial Intelligence 102: 104278. .

Assran, Mahmoud, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. 2023. "Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15619–29. .

Chen, Hanlin, Renyuan Luo, and Yiheng Feng. 2023. "Improving Autonomous Vehicle Mapping and Navigation in Work Zones Using Crowdsourcing Vehicle Trajectories." arXiv Preprint arXiv:2301.09194. .

Dehman, A., and B. Farooq. 2021. "Are Work Zones and Connected Automated Vehicles Ready for a Harmonious Coexistence? A Scoping Review and Research Agenda." Transportation Research Part C: Emerging Technologies 133: 103422. .

Gezici, Sinan, Zhi Tian, Georgios B. Giannakis, Hisashi Kobayashi, Andreas F. Molisch, H. Vincent Poor, and Zafer Sahinoglu. 2005. "Localization via Ultra-Wideband Radios: A Look at Positioning Aspects for Future Sensor Networks." IEEE Signal Processing Magazine 22 (4): 70–84. .

Ghosh, Anurag, Shen Zheng, Robert Tamburo, Khiem Vuong, Juan Alvarez-Padilla, Hailiang Zhu, Michael Cardei, Nicholas Dunn, Christoph Mertz, and Srinivasa G Narasimhan. 2025. "Roadwork: A Dataset and Benchmark for Learning to Recognize, Observe, Analyze and Drive Through Work Zones." In Proceedings of the IEEE/CVF International Conference on Computer Vision, 6132–42.

Habib, Ayman, Yun-Jou Lin, Radhika Ravi, Tamer Shamseldin, and Magdy Elbahnasawy. 2018. "LiDAR-Based Mobile Mapping System for Lane Width Estimation in Work Zones." FHWA/IN/JTRP-2018/10. Joint Transportation Research Program, Purdue University. .

Lee, Juho, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. 2019. "Set Transformer: A Framework for Attention-Based Permutation-Invariant Neural Networks." In Proceedings of the 36th International Conference on Machine Learning, 97:3744–53. Proceedings of Machine Learning Research. PMLR. .

Li, Hangyu, Xiaotong Sun, Chenglin Zhuang, and Xiaopeng Li. 2025. "On the Robotic Uncertainty of Fully Autonomous Traffic: From Stochastic Car-Following to Mobility–Safety Trade-Offs." Transportation Research Part C: Emerging Technologies 178: 105254.

Li, Qiyue, Zhong Wu, Jie Li, Wei Sun, and Jianping Wang. 2015. "A Novel Adaptive Kalman Filter Based NLOS Error Mitigation Algorithm." IFAC-PapersOnLine 48 (28): 1118–23. .

Li, Yuxiao, Santiago Mazuelas, and Yuan Shen. 2023. "A Semi-Supervised Learning Approach for Ranging Error Mitigation Based on UWB Waveform." arXiv Preprint arXiv:2305.18208. .

Liu, Jiaxi, Bolin Gao, Wei Zhong, Yanbo Lu, and Shuo Han. 2024. "Adaptive Optimization Strategy and Evaluation of Vehicle-Road Collaborative Perception Algorithm in Real-Time Settings." Computers and Electrical Engineering 120: 109785.

Maalek, Rami, and Farid Sadeghpour. 2016. "Accuracy Assessment of Ultra-Wide Band Technology in Locating Dynamic Resources in Indoor Scenarios." Automation in Construction 63: 12–26. .

Niu, Ziang, Huizhen Yang, Lei Zhou, Mohamed Farag Taha, Yong He, and Zhengjun Qiu. 2023. "Deep Learning-Based Ranging Error Mitigation Method for UWB Localization System in Greenhouse." Computers and Electronics in Agriculture 205: 107573.

Ochoa-de-Eribe-Landaberea, Aitor, Leticia Zamora-Cadenas, and Igone Velez. 2024. "Untethered Ultra-Wideband-Based Real-Time Locating System for Road-Worker Safety." Sensors 24 (8): 2391. .

Seo, Sungchul, Donghui Chen, Kwangcheol Kim, Kyubyung Kang, Dan Koo, Myungjin Chae, and Hyung Keun Park. 2022. "Temporary Traffic Control Device Detection for Road Construction Projects Using Deep Learning Application." In Construction Research Congress 2022, 392–401. American Society of Civil Engineers (ASCE). .

Shi, Weijing, and Ragunathan Raj Rajkumar. 2021. "Work Zone Detection for Autonomous Vehicles." In 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 1585–91. IEEE.

U.S. Department of Transportation. 2024. “Work Zone Data Exchange (WZDx).” .

Volpi, Andrea, Letizia Tebaldi, Guido Matrella, Roberto Montanari, and Eleonora Bottani. 2023. “Low-Cost UWB Based Real-Time Locating System: Development, Lab Test, Industrial Implementation and Economic Assessment.” Sensors 23 (3): 1124. .

Wang, Fang, Hai Tang, and Jialei Chen. 2023. “Survey on NLOS Identification and Error Mitigation for UWB Indoor Positioning.” Electronics 12 (7): 1678. .

Wang, Tianyu, Ke Hu, Zhenjie Li, Kai Lin, Jinzhao Wang, and Yuan Shen. 2021. “A Semi-Supervised Learning Approach for UWB Ranging Error Mitigation.” IEEE Wireless Communications Letters 10 (3): 688–91. .

Yang, Hongchao, Yunjia Wang, Shenglei Xu, Jingxue Bi, Haonan Jia, and Cheekiat Seow. 2024. “Ultra-Wideband Ranging Error Mitigation with Novel Channel Impulse Response Feature Parameters and Two-Step Non-Line-of-Sight Identification.” Sensors 24 (5): 1703.

Zaheer, Manzil, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. 2017. “Deep Sets.” In Advances in Neural Information Processing Systems, 30:3391–401.

Zuo, Fan, Jianbo Gao, Kaan Ozbay, Zia Bian, and Zhiqiang Zhang. 2023. “Urban Work Zone Detection and Sizing: A Data-Centric Training and Topology-Based Inference Approach.” In 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 3235–40. IEEE.

Gan, Rui, Junwei You, Yang Cheng, Xingyou Yang, Jiaxi Liu, Steven T. Parker, and Bin Ran. 2026. “A Unified LLM Framework for Work Zone Data Provision: Enhancement for Work Zone Data Exchange Feeds.” Manuscript prepared for TRB 2026.

Wu, Keshu, Yang Cheng, Junwei You, Rui Gan, Xuyan Zhao, Jingwen Zhu, Bin Ran, and Steven T. Parker. 2025. “Prototype of a Cost-Effective Roadside Device System for Vehicle Positioning in Cooperative Autonomous Driving.” Manuscript prepared for TRB 2025.

U.S. Department of Transportation. 2024. “Work Zone Data Exchange (WZDx) Specification and Work Zone Data Resources.”

Appendix A. Performance Indicator Synopsis

Table 17. Performance indicator synopsis.

Indicator area	Item	Synopsis
Part I - Research output	Final report	1 integrated final report combining CRM prototype evaluation, V2I/UWB denoising and mapping, and CRM-enhanced WZDx/LLM work zone data provision.
Part I - Research output	Publications / manuscripts	3 manuscript/report outputs documented: CRM prototype localization; V2I work zone geometry reconstruction with pose-conditioned UWB denoising; unified LLM framework for WZDx feed enhancement.
Part I - Data product	Experimental dataset	Dataset documentation package prepared for CRM static/dynamic experiments, V2I UWB field/simulation data, and WZDx/LLM case-study inputs/outputs.
Part I - Methodology	New methodology / technique	UWB triangulation and smoothing; pose-conditioned permutation-equivariant UWB denoising; cone localization and work zone polygon reconstruction; CRM-enhanced WZDx/LLM prompt design.
Part I - Software/model	Analytical model / code	Triangulation, denoising, calibration, downstream solver, WZDx/LLM prompt, and validation scripts documented; release status to be confirmed.
Part I - Equipment/instrument	CRM/UWB experimental setup	Cone-mounted CRM/UWB anchors, tag CRM/mobile unit, vehicle-mounted UWB tag, RTK/INS reference equipment, and local processing setup.
Part II - Outcome	Knowledge transfer	Improves understanding of low-cost infrastructure support for CAV localization and work zone data provision.
Part II - Workforce development	Students/researchers trained	Graduate/student involvement in hardware, software, data collection, analysis, AI/ML modeling, WZDx data, and report preparation; names/counts to be completed.
Part II - Impact	Safety/mobility potential	Potential to improve CAV work zone navigation by providing more accurate, timely, and interpretable infrastructure-derived geometry and lane guidance.
Part II - Impact	Cost/deployment potential	Low-cost CRM/UWB deployment may reduce the barrier to temporary work zone digital infrastructure and support targeted deployment at high-value locations.

Appendix B. Proposal Task Compliance Matrix

The matrix below repeats the project-task alignment in acceptance language, showing how each proposal task is represented in the final report and dataset package.

Table 18. Proposal task compliance matrix.

Task	Final report evidence	Dataset or output evidence
Task 1. System Assembly	Chapter 7 documents CRM prototype hardware, anchor/tag setup, UWB ranging mechanism, and four-anchor architecture; Chapter 3 documents field UWB-RSU setup.	Equipment output category, figures, and CRM prototype raw/processed folders.
Task 2. Distance Measurement Calibration	Chapter 7 reports static/dynamic CRM distance error distributions; Chapters 3-6 report UWB observation model, calibration, and raw-vs-denoised range metrics.	Calibration parameters, UWB range fields, and CRM error-result files in data dictionary.
Task 3. Vehicle Localization Performance Evaluation	Chapter 7 reports two-dimensional static/dynamic tag positioning; Chapters 5-6 report cone localization and Anchor Error_w metrics.	Tag/pose streams, anchor positions, estimated coordinates, denoised ranges, and localization result tables.
Task 4. Work Zone Boundary Mapping with CRM Integration	Chapters 3-6 document cone localization, polygon reconstruction, IoU, Hausdorff distance, and progressive reconstruction; Chapter 8 uses CRM cone coordinates for corridor refinement.	Mapped anchor fields, polygon outputs, CRM cone coordinate GeoJSON, and refined polyline outputs.
Task 5. WZDx Data Feed Prototype	Chapter 8 provides the CRM-enhanced WZDx/LLM workflow and case studies that translate WZDx descriptions/lane lists and CRM geometry into refined AV-ready work zone outputs.	WZDx case-study folders, prompt templates, recommended lanes, navigation instructions, and refined GeoJSON polylines.
Task 6. Report and Technology Transfer	Final report, manuscripts, recommendations, output/outcome/impact documentation, performance indicators, and dataset readiness package.	Dataset package, code/prompt archive, and future presentation/publication entries.

Appendix C. Experimental Dataset Data Dictionary and Acceptance Checklist

Table 19. Dataset data dictionary.

Field name	Type / unit	Definition
episode_id	string	Unique identifier for field or simulation episode.
dataset_split	categorical	train, validation, test, or field_diagnostic.
timestamp_s	float	Time stamp in seconds after episode start.
bin_id	integer	0.10 s aggregation bin identifier for field processed data.
tag_x_m, tag_y_m	float	2D vehicle-mounted UWB tag position in the local coordinate frame.
tag_z_m	float	Tag height above ground, if available.
gnss_x_m, gnss_y_m	float	RTK/INS reference position in the local coordinate frame.
anchor_id	string/integer	Unique UWB-RSU/CRM anchor identifier.
anchor_x_m, anchor_y_m	float	Anchor/cone position in the local coordinate frame.
observed_range_m	float	Raw or calibrated UWB observed range.
gt_range_m	float	Geometry-derived or simulated ground-truth distance.
valid_mask	0/1	Whether anchor measurement is valid/present at the time step.
nlos_indicator	0/1/NA	Simulation NLOS label or field unavailable indicator.
denoised_range_m	float	Corrected range output from a denoising method.
method	string	Raw, Kalman, MLP, PoseMLP, PoseKalman, or Proposed.
anchor_error_w_m	float	Weighted mean Euclidean cone-position error.
polygon_iou	float	Intersection-over-union of reconstructed and reference work zone polygons.
hausdorff_m	float	Symmetric Hausdorff distance between reconstructed and reference polygons.
calibration_version	string	Identifier for input/output calibration applied.
source_file	string	Original raw or processed file name.
crm_anchor_pair	string	Anchor pair used for two-anchor or four-anchor triangulation.
estimated_x_m, estimated_y_m	float / meter	Raw two-dimensional tag coordinate estimated from CRM triangulation.
smoothed_x_m, smoothed_y_m	float / meter	Time-smoothed coordinate estimate used for dynamic trajectory analysis.
anomaly_flag	0/1	Whether a CRM coordinate estimate exceeds the selected anomaly threshold.
wzdx_event_id	string	Identifier for the WZDx work zone event or case-study record.
wzdx_description	string	Natural-language WZDx lane closure description used as LLM input.
lane_order, lane_type, lane_status	integer/string	Structured WZDx lane-list fields indicating lane ordering, type, and open/closed status.
wzdx_polyline_geojson	GeoJSON LineString	Original rough WZDx work zone polyline supplied to the LLM.
crm_cone_coordinates_geojson	GeoJSON point list	CRM-derived cone or barrel coordinates used as high-resolution physical geometry.
recommended_lanes	list/integer	LLM-produced list of open lanes recommended for travel.
navigation_instruction	string	Human-readable LLM navigation instruction for the work zone.
refined_polyline_geojson	GeoJSON LineString	LLM-produced centerline polyline for the safe open corridor.
prompt_template_version	string	Version identifier for the WZDx/LLM prompt template.
llm_model_version	string	LLM or reasoning engine version used for case-study output generation.
validation_status	categorical	Pass/fail/review-needed label from schema, lane-status, and geometry checks.

- Raw data files have been submitted to Dryad with DOI: https://doi.org/10.5061/dryad.xgxd254xf

- Processed pose-range-mask sequences are reproducible from raw logs.
- Coordinate frame and units are explicitly defined.
- Anchor IDs are consistent within episodes and mapped to geometry files.
- Train/validation/test splits are episode-level and prevent trajectory leakage.
- Field calibration parameters are documented separately from final evaluation metrics.
- All result tables in the report are included as CSV files.
- Manifest includes checksums and version numbers for all data files.
- README includes contact, citation, and data-use restrictions.