Data Engineering and Applications

Research Paper

Data-Driven Port Perception Sequence Modelling with Graphormer-KAN for Autonomous Surface Vessel Berthing

Hamdan Al-Suwaidi[1], Ahmed Al-Mansoori[1,*], Rashid Al-Jabri[1] and Zayed Al-Tamimi[1]

[1] School of Engineering, Computer Science Department, Khalifa University, Abu Dhabi, PO Box 127788, United Arab Emirates
*Corresponding author: ahmed.man@kustar.ac.ae

Abstract. Close-quarters berthing is still a difficult autonomy problem because useful control decisions need to be derived from asynchronous port observations, and hydrodynamic response changes rapidly near the quay. This paper proposes a port-perception-sequence-driven Graphormer-KAN model that converts synchronized Lidar, radar, camera, positioning and vessel-state measurements into dynamic heterogeneous graphs. Spatial relation biases in the Graphormer are berth topology, visibility and collision relevance; a Kolmogorov-Arnold Network decoder generates distributional motion increments and actuator references. Predicted Uncertainty Expands Clearance Constraints in a Receding-Horizon Safety Controller. A model of a 9.2m twin-propeller vessel was used to study 2,400 random berthing episodes under four berth layouts, five disturbance levels, three payload states and structured sensor degradation. The proposed model had a successful berthing rate of 94.8%, a median terminal-position error of 0.19 m, a median heading error of 1.42°, and a fifth-percentile minimum clearance of 0.46 m. Success increased by 6.9 percentage points and collision incidence decreased from 5.7% to 1.8% compared with the strongest transformer baseline. Calibration error was 0.031, and the median inference time on an embedded graphics processor was still 18.6ms. Based on ablation experiments, relation-biased attention, spline-based KAN decoding and uncertainty-conditioned margins are all beneficial. According to the above results, graph-structured perception sequences can perform accurate, calibrated and computationally feasible autonomous berthing in heterogeneous port environments.

Keywords: *Autonomous Vessel Berthing, Port Perception Sequence, Graphormer-KAN, Uncertainty Calibration, Safety-Constrained Control*

Received on 25 February 2023, Accepted on 06 June 2023, Published on 19 June 2023

Introduction

As commercial and service ports gradually introduce small autonomous surface vessels (ASVs), they will increasingly operate near crewed vessels, fixed structures and change rapidly with shifting water boundaries. Berthing is the last manoeuvre in the terminal approach, and small errors in the state estimation can lead to contact because a large amount of momentum needs to be dissipated at a low speed. At present, the stacks for marine autonomy integrate global navigation satellite systems (GNSS), inertial sensing, radar, lidar and vision to achieve full coverage [1]. Close-range LiDAR can accurately resolve the geometry of a quay, and radar is unaffected by rain and fog [2]. Camera-based semantic cues support fender and bollard recognition but are sensitive to illumination [3]. Multi-sensor fusion thus enhances observability, but asynchronous sampling and partial occlusion are still present; nevertheless, the recent measurement sequence is more informative than any single frame [4].

Current autonomous berthing controllers are generally in the form of gain-scheduled feedback or model-predictive/learning-based policies. Model-predictive control can explicitly show actuator limits and geometric constraints [5]. However, the performance will drop if the prediction model fails to consider shallow-water interaction, windage and propulsor nonlinearities [6]. Deep sequence models can learn residual dynamics and manipulation patterns from data, but flat feature vectors cannot retain the constantly changing relationships between the hull, dock, fenders, and nearby ships [7]. Graph-neural network representations are a natural choice as they capture entities and pairwise relationships, but traditional message passing may lose long-distance, safety-critical information and remain poorly calibrated under distribution shift [8]. The defect is particularly severe in the final stages; otherwise, the controller will not be able to determine whether the narrow prediction gap is reliable.

This paper introduces a Port-Perception-Sequence-driven Graphormer-KAN architecture to address the problem of port-scene perception and uncertainty-aware close-quarters control. Each synchronous observation is represented as a heterogeneous graph, and its nodes include the controlled vessel, berth geometry, fenders, detected obstacles and environmental estimates. A Temporal Graphormer creates interaction-aware tokens by using relation, distance, visibility and collision-relevance biases. A Kolmogorov-Arnold Network (KAN) decoder is used to map the tokens to distributional motion increments and actuator references via learnable univariate spline functions, thus achieving localized nonlinear response without an opaque multilayer perceptron. Decode the covariance, and then set an adaptive clearance margin in a receding-horizon controller using it. Accuracy, calibration, safety, computation, ablation and cross-portability are also used as evaluation criteria instead of trajectory error.

Related Work

Perception-Based Autonomous Berthing

Research on autonomous docking is gradually shifting from instrument-equipped ports and predefined approach trajectories to onboard perception and constrained feedback. Nonlinear model predictive controllers can set upper limits on terminal attitude, speed, and actuator constraints [9]. Data-driven maneuvering models reduce the difficulty of hydrodynamic identification, especially in the coupling of low-speed roll and yaw [10]. Reinforcement-learning policies have also shown a direct mapping from vessel states to thruster commands, but their safety during distribution is highly dependent on training coverage and auxiliary constraints [11]. All of these ways generally reduce perception to a fixed-occupancy vector or a small number of distances. Such compression fails to show which quay segment produced the return, how an obstacle relates to the intended corridor, and whether a missing measurement is due to visibility or absence.

Sequence-aware fusion is required due to latency, intermittent returns and motion-dependent occlusion in a harbor observation. A feasible berthing system should have different measurement frequencies, propagate uncertainty through an estimator, and distinguish between fixed facilities and moving objects. Many reported controllers assume a state vector that has already been cleaned and localized. That assumption ignores perception errors and control evaluation, thus exaggerating closed-loop robustness. The current formulation maintains the identity of entities, observation age and geometric relations in all graph snapshots, so a degraded sensing condition changes both the node confidence and the attention bias given to the predictor.

Graph Sequence Models and Nonlinear Decoders

GNNs can explicitly model the structure of relations, and have been applied to traffic forecasting, interaction prediction and multi-agent motion modeling [12]. Transformer variants extend the receptive field by performing global attention, and graph-specific structural encodings prevent attention from treating unrelated node pairs as equivalent [13]. Graphormer shows that shortest-path and centrality encodings can add topology to transformer attention [14]. Port scenes, however, do not have a simple shortest-path structure; Euclidean clearance, line-of-sight, relative bearing and time-to-collision are all different physical quantities that need to be considered separately.

KANs replace fixed scalar activations on edges with learnable univariate functions to offer a compact form of multivariate nonlinear mapping approximation [15]. Spline structures are suitable for low-speed marine motion because the control effect varies locally with speed, angle and environmental load. However, KAN cannot organize variable cardinality port scenarios, while graph transformers typically end with traditional multi-layer perceptrons, which require a large number of parameters to capture the responses of local actuators. The roles of the above components have been set up as follows: Graphormer addresses the problem of temporal relational context, KAN conducts nonlinear distribution decoding, and a constrained controller maps calibrated predictions into executable instructions.

Dynamic Port-Scene Graph Representation

Sequence Synchronization and Node Construction

The Perception front-end runs at a 10 Hz control rate. Lidar clusters, radar tracks, camera detections, automatic identification messages, inertial states and actuator feedback are all converted to a berth-fixed coordinate system. Measurements are propagated to the control timestamp with their reported covariances and are not considered simultaneous. Each graph snapshot contains an ego-vessel node, polyline-derived quay nodes, fender nodes, tracked-obstacle nodes, and an environmental node. Node features are position, velocity, heading, dimensions, semantic confidence, observation age, visibility and covariance descriptors. Quay polylines are resampled at 1.5 m intervals in the vicinity of the planned contact area and more coarsely elsewhere to maintain local shape fidelity while keeping the token count reasonable. At synchronized time t, node i is represented by a compact state vector containing geometric, kinematic, semantic, and confidence information:

$$\boldsymbol{x}_i^t = [\boldsymbol{p}_i^t, \boldsymbol{v}_i^t, \psi_i^t, \boldsymbol{d}_i, \boldsymbol{c}_i, a_i^t, \boldsymbol{q}_i^t] \qquad \text{Eq.(1)}$$

Here, $\mathbf{p}_i^t$ and $\mathbf{v}_i^t$ denote position and velocity, ψ_i^t is heading, $\mathbf{d}_i$ describes physical dimensions, $\mathbf{c}_i$ contains semantic attributes, a_i^t is observation age, and $\mathbf{q}_i^t$ summarizes measurement quality. Missing channels are masked rather than zero-imputed as physical values. A sliding window of 12 snapshots contains 1.2 seconds of past data. All observations are shifted from their sensor times τ to the common control time. Increase the corresponding covariance to maintain a constant uncertainty introduced by temporal alignment:

$$\begin{aligned} \widetilde{\boldsymbol{p}}_i^t &= \boldsymbol{p}_i^\tau + \Delta t \boldsymbol{v}_i^\tau \\ \widetilde{\boldsymbol{\Sigma}}_i^t &= \boldsymbol{\Sigma}_i^\tau + \Delta t \boldsymbol{Q}_i \end{aligned} \qquad \text{Eq.(2)}$$

where $\Delta t = t - \tau$, and $\mathbf{Q}_i$ is the propagation-noise matrix.

Edges are directed because the operational relevance of a quay segment to the vessel differs from that of the vessel to the quay. Four edge families represent physical adjacency, visibility, kinematic interaction, and planned-corridor membership. Their continuous attributes capture range r_{ij}^t, relative bearing β_{ij}^t, closing speed Δv_{ij}^t, time to collision, and line of sight:

$$\boldsymbol{e}_{ij}^t = \left[r_{ij}^t, \beta_{ij}^t, \Delta v_{ij}^t, TTC_{ij}^t, \ell_{ij}^t\right] \qquad \text{Eq.(3)}$$

The entire perception-to-actuation workflow is shown in Figure 1 below. Asynchronous port observation first undergoes coordinate transformation, covariance-aware synchronization, and dynamic graph construction. Graph sequence is then processed by the Graphormer-KAN model, and finally, the safety controller converts the predicted vessel motion and uncertainty into executable actuator commands.

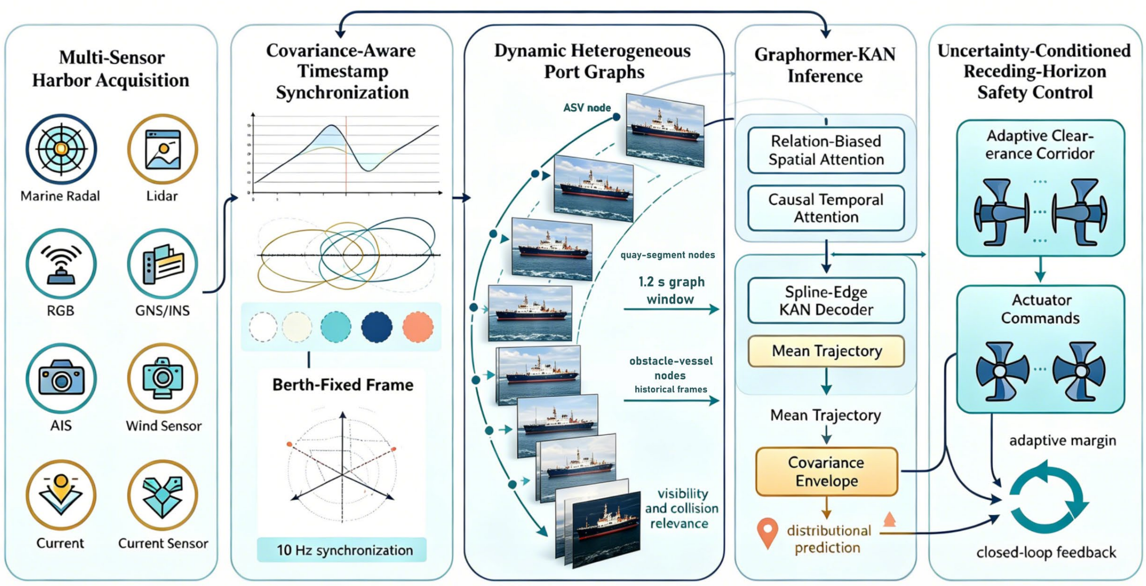

Figure 1. Port-aware sequence workflow.

Relation Encoding and Temporal Packaging

Note that deviations should prioritize topological relationships over physical proximity. For each ordered node pair, relation type, clipped graph distance, visibility state and collision-relevance class are embedded separately. Metric distance is presented as a radial basis expansion. The resulting attention bias is shown as follows:

$$b_{ij}^{t} = \boldsymbol{e}_{ij}^{r} + \boldsymbol{e}_{ij}^{g} + \boldsymbol{e}_{ij}^{v} + \boldsymbol{e}_{ij}^{c} + \boldsymbol{w}_{d}^{\top}\phi\left(r_{ij}^{t}\right) \quad \text{Eq.(4)}$$

The first four terms are the learned embeddings of relation type, graph distance and visibility, and collision relevance. The last term is a continuous-metric distance function.

Snapshots are connected by identity edges for persistent objects and soft association edges for uncertain tracks. Temporal age is relative to the most recent snapshot; thus, old evidence is reduced in weight but not deleted immediately. The first token is the predicted node state plus structural and temporal encodings:

$$\boldsymbol{h}_{i}^{t,0} = \boldsymbol{W}_{x}\boldsymbol{x}_{i}^{t} + \boldsymbol{e}_{i}^{deg} + \boldsymbol{e}_{t}^{time} \quad \text{Eq.(5)}$$

The degree encoding $\mathbf{e}_{i}^{\mathrm{deg}}$ combines the in-degree and out-degree information of node i, while $\mathbf{e}_{t}^{\mathrm{time}}$ identifies the relative temporal position of the snapshot.

Graph size is limited to 96 nodes in a snapshot. If the number of detections exceeds this budget, then all quays and fender nodes within the stopping envelope are kept, and dynamic obstacles are ranked based on predicted collision relevance. The Policy keeps the local geometry required by the final alignment. Normalization parameters are only fitted on the training set. Data augmentation perturbs timestamps, masks sensors in contiguous bursts, and displaces quay points within their covariance envelopes to make the model learn temporal consistency and measurement uncertainty instead of depending on artificially clean data.

Uncertainty-Aware Graphormer-KAN Berthing Control

Spatial-Temporal Graphormer Encoder

Each graph snapshot is first processed by relation-biased multi-head self-attention. Query-key compatibility is extended by adding a port-specific bias introduced in Section 3, and global attention can still retain geometric and visibility information. For attention head m_{r} the normalized influence of node j on node i is

$$\alpha_{ij}^{m,t} = softmax_{j}\left(\frac{\boldsymbol{q}_{i}^{\top}\boldsymbol{k}_{j}}{\sqrt{d_{h}}} + b_{ij}^{m,t}\right) \quad \text{Eq.(6)}$$

where d_{h} is the feature dimension of one attention head.

The attended message is introduced through a gated residual path. This gate suppresses abrupt representation changes when track identities switch or a sensor becomes temporarily unavailable:

$$\boldsymbol{h}_{i}' = LN(\boldsymbol{h}_{i} + \boldsymbol{g}_{i} \odot \boldsymbol{m}_{i}) \quad \text{Eq.(7)}$$

Here, $\mathbf{m}_{i}$ is the attention-weighted message and $\mathbf{g}_{i}$ is a sigmoid gate learned from the current node and message states.

Temporal attention is then applied to the ego token and the pooled safety tokens of the historical snapshots. Causal Masking does not access future observations. The sequence representation is given by the weighted sum:

$$\boldsymbol{s}_{T} = \sum_{t=T-L+1}^{T} \omega_{t}\boldsymbol{z}_{E}^{t} \quad \text{Eq.(8)}$$

The coefficient ω_{t} is learned from token relevance, observation confidence, and temporal age. Consequently, recent information usually receives a larger weight, while an older but high-confidence observation can remain influential during short sensor outages.

The four spaces and the two times, eight attention heads and a hidden size of 128 make up the encoder. Relation biases are shared by the spatial layers for memory saving. A bounded token budget and causal construction can be used for online inference while retaining enough historical information to identify the gradual sway accumulation that may not be obvious in a single observation frame.

KAN Distributional Decoder

The KAN decoder represents each output as a sum of learned univariate transformations. Unlike a dense decoder that mixes all latent coordinates through fixed activation functions, the KAN can allocate spline knots around low-speed regimes in which actuator response changes sharply:

$$y_q = \sum_{p=1}^{d} \Phi_{qp}(s_p) \quad \text{Eq.(9)}$$

Each transformation combines a smooth base function with a cubic B-spline expansion:

$$\Phi_{qp}(s) = w_{qp} SiLU(s) + \sum_{k} c_{qpk} B_k(s) \quad \text{Eq.(10)}$$

The decoder predicts surge, sway, yaw-rate, pose, and actuator increments over a 20-step horizon. It also produces a lower-triangular covariance factor whose diagonal terms are constrained to be positive. Let $\boldsymbol{\delta}_k = \mathbf{y}_k - \boldsymbol{\mu}_k$ denote the prediction residual. Trajectory supervision uses the Gaussian negative log-likelihood

$$\mathcal{L}_{nll} = \frac{1}{2} \sum_{k} \left(\boldsymbol{\delta}_k^{\top} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k + log|\boldsymbol{\Sigma}_k|\right) \quad \text{Eq.(11)}$$

A collision-aware term penalizes trajectories that enter the protected boundary. The complete objective combines distributional accuracy, terminal-pose accuracy, collision exposure, and command smoothness:

$$\mathcal{L} = \mathcal{L}_{nll} + \lambda_p \mathcal{L}_p + \lambda_\psi \mathcal{L}_\psi + \lambda_c \mathcal{L}_c + \lambda_u \mathcal{L}_u \quad \text{Eq.(12)}$$

The four auxiliary terms are the terminal-position error, the terminal-heading error, collision exposure and actuator variation. The weights are chosen from the validation set and not from the test cases.

The spline grid is uniformly distributed in the normalized latent space and is slightly offset during training according to the quantile function. Curvature regularization avoids oscillatory edge functions, while node sparsification removes small coefficients before deployment. Thus, the decoder is more stable when the environment falls outside the operating range shown in the training data.

Uncertainty-Conditioned Receding-Horizon Control

The controller obtains the decoded average trajectory, covariance, and feedforward actuator sequence. By extending the nominal boundary margin along the local boundary normal direction, the confidence is converted into geometric constraints.

$$m_k = m_0 + \kappa \sigma_k^n \quad \text{Eq.(13)}$$

The directional standard deviation is obtained from the predicted position covariance:

$$\sigma_k^n = \sqrt{\boldsymbol{n}_k^{\top} \boldsymbol{\Sigma}_{p,k} \boldsymbol{n}_k} \quad \text{Eq.(14)}$$

At every control step, a quadratic program adjusts the feedforward actuator sequence. The objective balances trajectory tracking, command correction, and control smoothness:

$$J = \sum_{k} \left(\left\| \boldsymbol{x}_k - \boldsymbol{x}_k^{ref} \right\|_{\boldsymbol{Q}}^2 + \| \boldsymbol{u}_k - \overline{\boldsymbol{u}}_k \|_{\boldsymbol{R}}^2 + \| \Delta \boldsymbol{u}_k \|_{\boldsymbol{S}}^2 \right) \quad \text{Eq.(15)}$$

The command sequence is selected subject to the main safety constraint

$$\boldsymbol{u}^* = arg \min_{\boldsymbol{u}} J, d_k \geq m_k \quad \text{Eq.(16)}$$

Actuator magnitude, actuator rate and terminal-velocity constraints are all imposed simultaneously by the optimizer.

Due to the possibility that imperfect predictive distributions may still be overly confident, a conformity measure is estimated from the normalized validation residuals. The desired residual quantile is

$$\kappa_{cal} = Q_{1-\alpha}(r) \quad \text{Eq.(17)}$$

The calibrated covariance used by the safety controller becomes

$$\widehat{\boldsymbol{\Sigma}}_k = \kappa_{cal}^2 \boldsymbol{\Sigma}_k \quad \text{Eq.(18)}$$

This post-processing stage targets coverage level $1 - \alpha$ without retraining the Graphormer-KAN predictor.

https://doi.org/10.64972/dea.2023.v2i2.3608d:105-118

Figure 2 shows how relation-biased spatial attention and causal temporal aggregation feed the spline-based KAN distributional decoder. Conformal Scaling then normalizes the predicted covariance before it is fed into the receding-horizon safety layer. Therefore, the architecture maintains a clear division of labor between relational awareness, distribution prediction, uncertainty calibration, and constrained control.

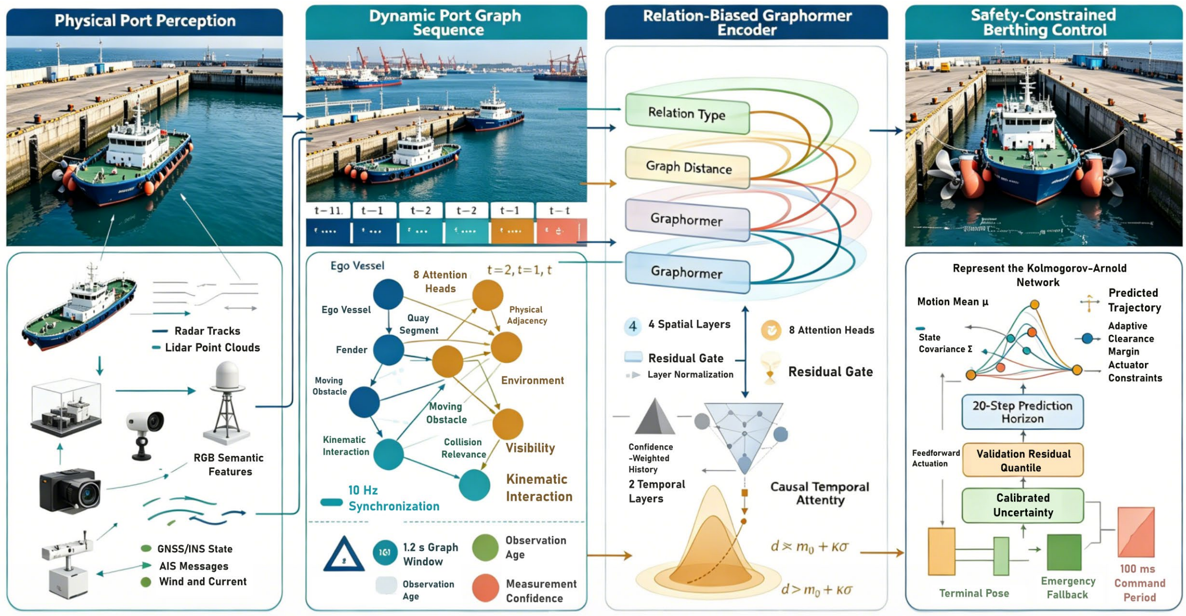

Figure 2. Detailed Graphormer-KAN architecture with relation-biased spatial attention.

Experimental Evaluation and Analysis

Experimental Protocol and Comparative Methods

Experiments used a high-fidelity planar simulation based on a 9.2 m x 3.1 m beam, twin-propeller workboat with a bow thruster. This set of actions consists of 2400 episodes: 1440 episodes for training, 360 episodes for validation and calibration, and 600 episodes for testing. Four berth arrangements included straight quay, recessed slip, angled face and narrow basin. Five environments had a constant current of 0.0-0.6 m/s, a gusting wind speed of 0-12 m/s, and wave-induced coloured disturbances. Three payload states altered mass and yaw inertia by up to 22%. Sensor conditions included normal operation, LiDAR dropout, camera glare, radar clutter, positioning bias and mixed degradation. The above factors are too far-reaching to be considered single conditions for concealing perception-control coupling [16].

The vessel was 18-32m away from the terminal pose and had heading offsets of -35° to 35°. Success required a terminal position error less than 0.45m, a heading error less than 4°, an absolute surge and sway speed less than 0.12m/s, no boundary contact, and completion within 140s. Each method received the same state estimate, graph candidate, training episode and controller limit. Baselines were a nonlinear model-predictive controller (NMPC), long short-term memory predictor plus MPC (LSTM-MPC), temporal convolutional network plus MPC (TCN-MPC), graph convolutional recurrent network plus MPC (GCRN-MPC), standard transformer plus MPC (Transformer-MPC), and Graphormer with a multilayer-perceptron decoder (Graphormer-MLP). Separating the practice of terminal accuracy from safety and runtime constitutes evaluation [17]. Bootstrap confidence intervals used 10,000 resamples at the episode level. Calibration was assessed using empirical coverage and expected calibration error, as the mean prediction error does not indicate overconfidence [18].

AdamW was used for 160 rounds of training; a maximum learning rate of 3×10^{-4}, cosine decay, a batch size of 12 sequences, and gradient clipping at 1.0 were set. Early stopping monitored the validation negative log-likelihood and closed-loop collision rate. The chosen model has 3.84 million parameters. Deployment measurements used an embedded GPU with a batch size of one and included graph assembly, inference, covariance scaling and safety-layer solution. Controller horizon is 2.0s, and the command period is 0.1s. Hyperparameters were set before opening the 600-episode test set.

Ground-truth vessel motion was interpolated to 100 Hz and then sampled at the sensor-specific rates. The radar model generated range-bearing returns with range-dependent noise, clutter bursts and missed detections behind the quay corner. Lidar scans included water-surface multipath and partial hull returns, and the camera model varied contrast, saturation and class confidence with the sun angle. Positioning bias followed a first-order process instead of independent white noise, and the error thus remained in the graph window. Wind loads are applied at the estimated pressure center, and there are spatial variations in the water flow near the dock face. Actuator dynamics include dead zones, rate limits, command delay and asymmetric thrust loss. These details are necessary to prevent the predictor from using the unrealistic independence of sensing and disturbance. Episode seeds, scenario definitions, initial conditions and sensor-fault schedules were saved prior to inference, and all baselines encountered the same event sequence. Training Trajectories were generated by combining expert NMPC, perturbed expert commands and manually specified recovery maneuvers. Samples within 0.8 m of the boundary were taken at double the normal rate so that the loss was not dominated by easy open-water motion.

Set the definitions for metrics at the episode and prediction levels. Position and heading errors were evaluated at all horizon steps, and terminal metrics were reported separately. The smallest distance to the oriented hull polygon was taken as the clearance. A contact event required polygon intersection for two consecutive integration steps, suppressing numerical grazing artifacts. Runtime statistics excluded model loading but included tensor preparation and quadratic-program solution. Six nominal coverages and equal weighting were used for the expected calibration error; collision reliability employed probability bins and presented the count in each bin to reveal sparse estimates. Generalisation tests excluded all observations from the three geometrically distinct ports in model fitting. The lightweight calibration performed at these ports only updated the normalized statistics, conformal scale, and final KAN spline coefficients; the Graphormer weights remained unchanged. Therefore, the transfer results will not be considered as unrecorded retraining experiments. Use paired episode seeds for statistical comparison, and calculate confidence intervals after obtaining all time steps in the episode to avoid pseudoreplication.

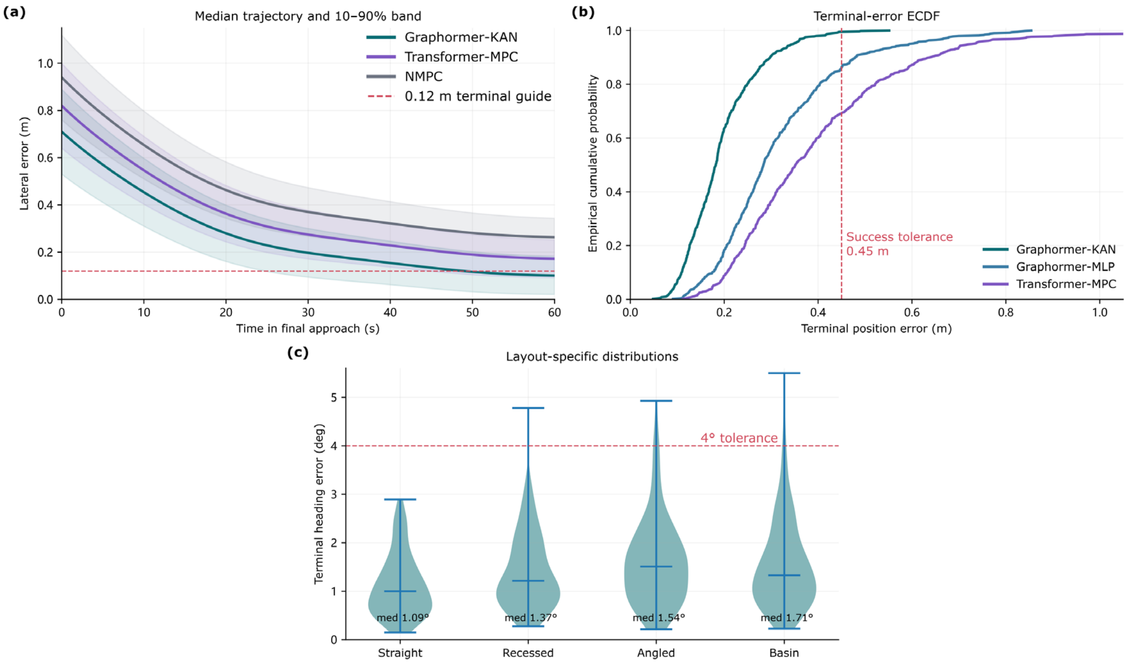

Figure 3. Accuracy across the berthing maneuver: (a) lateral-error trajectories with median and 10–90% bands; (b) empirical cumulative distributions of terminal position error; (c) terminal heading-error violins for four berth layouts.

Prediction, Calibration, and Closed-Loop Performance

The development and final distribution of berthing error are shown in Figure 3. In the last 60 seconds, as shown in Figure 3(a), the median lateral error of Graphormer-KAN drops from 0.71m to 0.08m; Transformer-MPC and NMPC reach 0.15m and 0.24m, respectively, and the shaded 10-90% band remains narrow after 35s. The empirical distributions in Figure 3(b) show that 90% of Graphormer-KAN trials are less than 0.38m, compared with 0.57m for Graphormer-MLP and 0.71m for Transformer-MPC. As shown in Figure 3(c) for the straight, recessed, angled and basin berths, the proposed heading-error medians are 1.09°, 1.37°, 1.54° and 1.71°, respectively, and the basin tail remains below 4° in 93.5% of the trials. Sequence models are known to be better at handling noisy motion derivatives with temporal context [19]. The maximum gain occurs when the ship is near the dock, and at this point, the graph bias connects the hull ends with a local part of the dock.

Robustness under environmental and sensing degradation is examined in Figure 4. The polar map in Figure 4(a) keeps success above 90% for every single-degradation condition and reaches 86.7% under the joint highest disturbance and mixed-sensor condition. Outcome decomposition in Figure 4(b) gives Graphormer-KAN 569 successes, 11 contacts, 8 timeouts, and 12 terminal-tolerance failures, compared with 528, 34, 14, and 24 for Graphormer-MLP. The clearance–current relationship in Figure 4(c) further shows that the fitted risk boundary separates 94.2% of contact events and that high-uncertainty commands shift toward clearances above 0.55 m. Robust marine control studies emphasize evaluating compound disturbances rather than isolated noise [20]. The joint degradation results indicate graceful decline rather than an abrupt failure threshold.

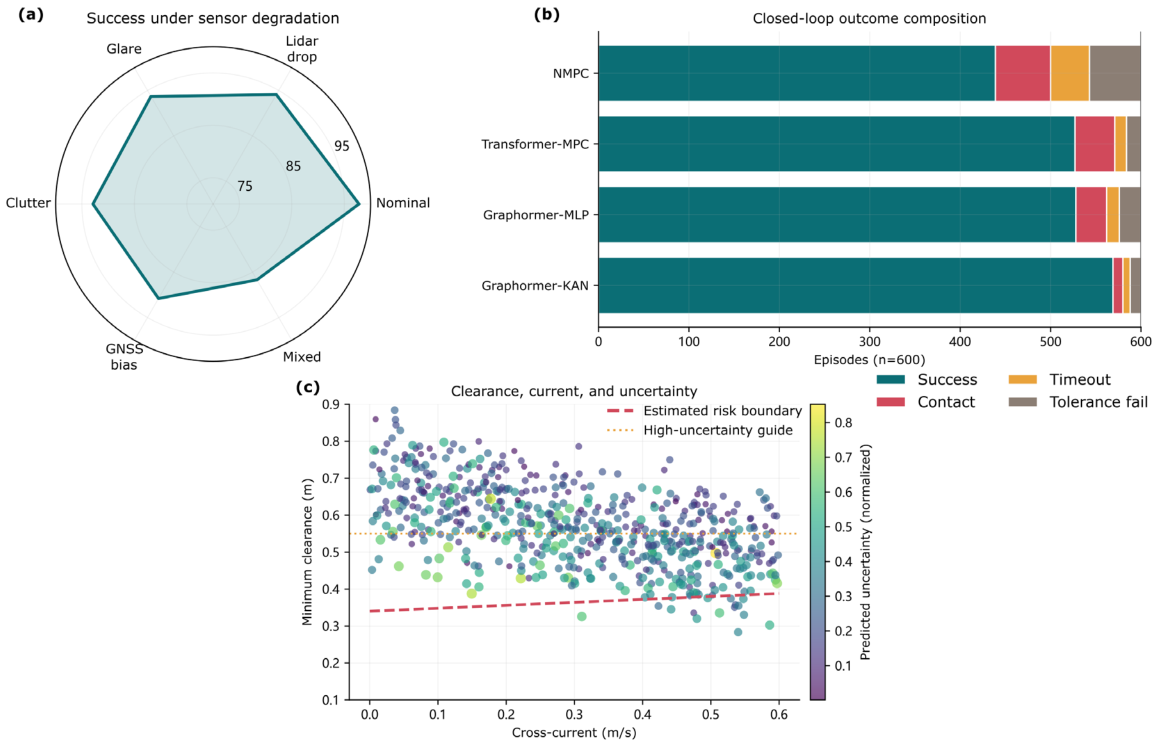

Figure 4. Robustness under environmental and sensing degradation: (a) polar success-rate map; (b) stacked outcome composition by method; (c) clearance–cross-current density scatter with uncertainty contours and risk boundary.

The Quality of the Predicted Uncertainty is shown in Figure 5. For nominal intervals of 50%, 70%, 80%, 90%, 95% and 99%, Figure 5(a) shows the post-calibration empirical coverage of 51.8%, 69.2%, 79.1%, 89.4%, 94.6% and 98.3%, and the expected calibration error has dropped from 0.087 to 0.031. As shown in Figure 5(b), the reliability view divides the collision probability into ten bins; among them, the 0.4-0.5 bin has 46 episodes and an observed event rate of 0.43, and most of the high-confidence safe bins are included. As shown in Figure 5(c), the hexagonal residual plot remains centered in both the lateral and yaw directions, with a correlation of 0.28 and a 95% confidence interval of ±0.31 meters and ±2.7°. Distributional prediction is only practical when the coverage aligns with the actual errors [21]. Thus, the calibrated covariance can be employed as a reasonable index for margin adjustment in lieu of an arbitrary safety multiplier.

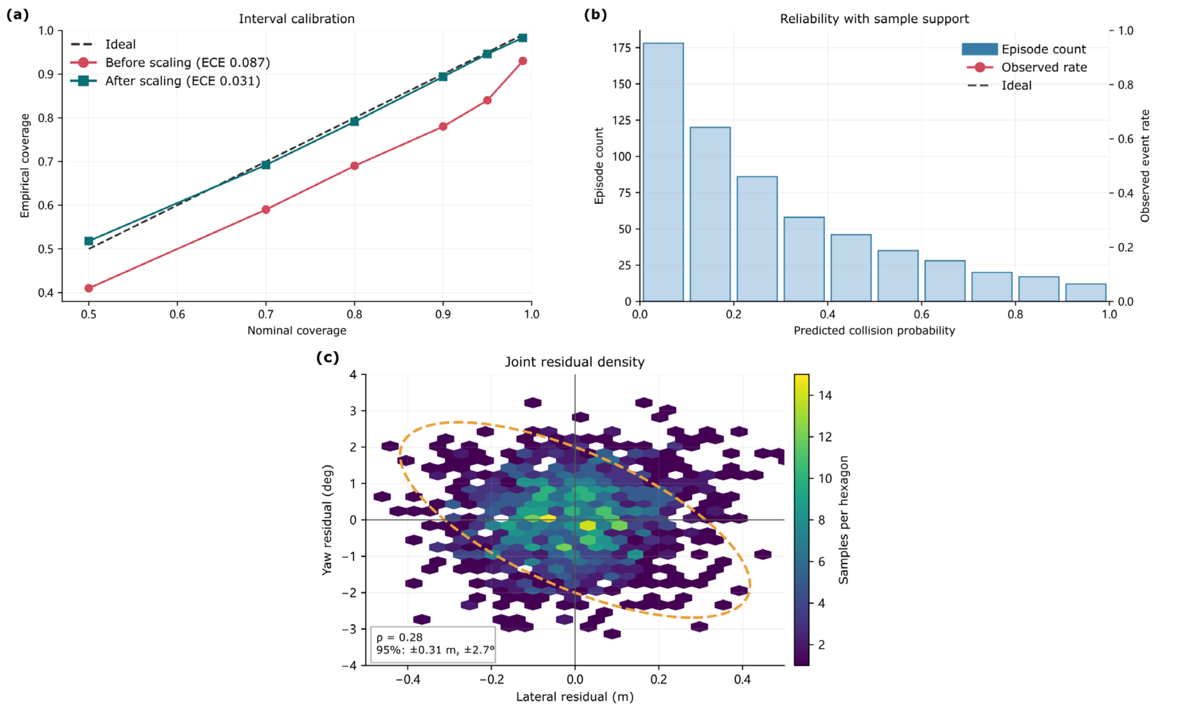

Figure 5. Predictive uncertainty and calibration: (a) coverage calibration curves before and after conformal scaling; (b) collision-probability reliability histogram with sample counts; (c) lateral–yaw residual hexbin density with covariance ellipse and marginal summaries.

The end-point indicators show the source of the total increase in success. Under the straight-quay condition, both Graphormer-KAN and Graphormer-MLP complete most of the episodes, but the KAN decoder reduces the 95th-percentile lateral error from 0.51 m to 0.39 m. Global attention is required for the recessed slip because quay segments on both sides of the entrance are simultaneously relevant; without relationship bias, late corrective reversals have increased by 31%. The angled face has the highest requirement for yaw-sway coordination. The proposed predictor anticipates the change in the sign of sway velocity about 1.4 seconds before the flat transformer does. Uncertainty adaptation is used for the narrow basin to enhance safety; the average commanded approach speed is reduced by 0.06 m/s when sensors degrade partially, and only differs by more than 0.01 m/s under normal sensing conditions. Therefore, a uniform slow-motion effect will not be achieved for the safety increase. The median completion time of the proposed system is 96.4s, that of Transformer-MPC is 95.1s, and that of NMPC is 103.7s. The small-time cost of the transformer is accompanied by a significantly lower contact rate; at the same time, the physics-only controller loses time due to repeated heading corrections.

Prediction Errors were also divided by Horizon. At 0.5 s, the lateral root-mean-square error is 0.041 m and the yaw-rate error is 0.19°/s; at 1.0 s, these have risen to 0.083 m and 0.37°/s, and at 2.0 s, they reached 0.168 m and 0.71°/s. Graphormer-MLP is almost the same at 0.5s, but it deviates at the longest horizon; its lateral error is 0.231m. KAN has advantages in episodes with a low surge speed and a large transverse environmental load, and they correspond to a local spline response. Error is not the same for all objects. With fewer than 25 graph nodes, the median terminal error is 0.17 m; for 25-55 nodes, it is 0.19 m; and beyond 55 nodes, it is 0.23 m. Inference time is increased by 4.8ms in all the above ranges but still meets the command budget. Track age has a clearer calibration effect: when the oldest dynamic-obstacle observation exceeds 0.6 s, the calibrated standard deviation increases by 42%, and the empirical 90% coverage is still 88.1%. Without observation-age features, the coverage in that subset is 73.6%. Therefore, the partitions show that the confidence reacts to recognisable sensing difficulty rather than only reflecting the mean trajectory error.

Ablation, Efficiency, and Generalization

Ablation experiments study the impact of a module. After removing the relation bias, the success rate was 90.5 per cent, down from 94.8 per cent, and the median position error was 0.27 metres instead of 0.19 metres. Replace the KAN decoder with a matched MLP to lower success to 88.0% and increase calibration error to 0.056.

Removing temporal history yields an 86.8% success rate, and a single snapshot cannot reliably infer the accumulation of slow drift. Set the clearance margin to 0.35 m; 91.2 per cent were successful, but the contacts reached 24. According to the above data, organised attention can improve interaction modelling when the topology contains task information [22]. They also show that the controller needs to be uncertain; reporting only the covariance does not increase safety.

Component effects and closed-loop dynamics are shown together in Figure 6. The box-and-swarm distributions in Figure 6(a) place the full-model interquartile range at 0.12–0.28 m, compared with 0.18–0.41 m without relation bias and 0.23–0.49 m without history. In the sway–yaw phase portrait of Figure 6(b), successful trajectories contract inside the prescribed 0.12 m/s and 2.5°/s envelope, whereas contact trials cross the upper-right boundary before 20 s. Contact-free survival in Figure 6(c) ends at 98.2% for the proposed method, 94.3% for Graphormer-MLP, 92.7% for Transformer-MPC, and 89.8% for NMPC. Ablation is especially important for hybrid learning-control systems because architectural gains can otherwise be confounded with conservative constraints [23].

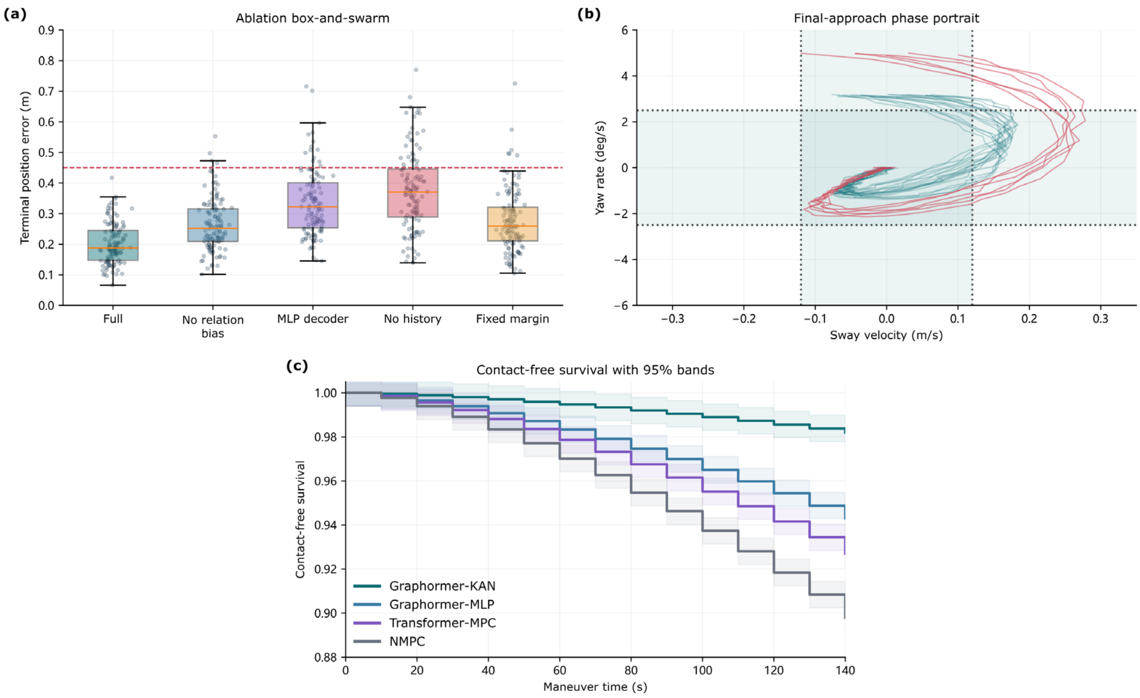

Figure 6. Component and dynamic diagnostics: (a) box-and-swarm terminal-error distributions for model ablations; (b) sway–yaw phase portrait with safe envelope and time coloring; (c) contact-free survival curves with confidence bands.

Deployability and cross-port transfer are considered in Figure 7. The Pareto map in Figure 7(a) places Graphormer-KAN at 94.8% success, 18.6 ms latency, and 3.84 million parameters; Graphormer-MLP reaches 88.0% at 16.9 ms and 4.11 million, while an eight-head larger transformer reaches 89.1% but requires 32.7 ms. Across nominal accuracy, robustness, calibration, efficiency, and unseen-port transfer, the radar profile in Figure 7(b) gives the proposed method normalized scores of 0.95, 0.91, 0.94, 0.88, and 0.87. The raincloud distributions in Figure 7(c) show median errors of 0.24 m, 0.27 m, and 0.31 m for three held-out ports before adaptation, falling to 0.20 m, 0.22 m, and 0.25 m after only 20 calibration episodes per port. Embedded inference must be assessed with the full preprocessing and control path rather than neural execution alone [24]. The latency remains below the 100 ms command period in 99.6% of calls, leaving capacity for tracking and health monitoring.

The overall closed-loop success rate is 94.8%, compared with 87.9% for Transformer-MPC, 88.0% for Graphormer-MLP, 83.7% for GCRN-MPC, 80.8% for TCN-MPC, 78.5% for LSTM-MPC, and 73.2% for NMPC. Collision incidence falls to 1.8%, and the fifth-percentile minimum clearance is 0.46 m. Nonparametric paired bootstrap tests give $p<0.01$ for the success and terminal-error differences against every baseline. Modern autonomous-vessel evaluation increasingly calls for scenario diversity and repeatable failure taxonomies [25]. Accordingly, contact, timeout, terminal-tolerance failure, perception divergence, and solver fallback were logged separately. Of the 11 proposed-model contacts, seven occurred during simultaneous positioning bias and severe cross-current, three followed persistent radar-lidar disagreement, and one followed a late dynamic-obstacle incursion. The small number of solver fallbacks did not produce contact because the previous feasible command and an emergency braking policy remained available.

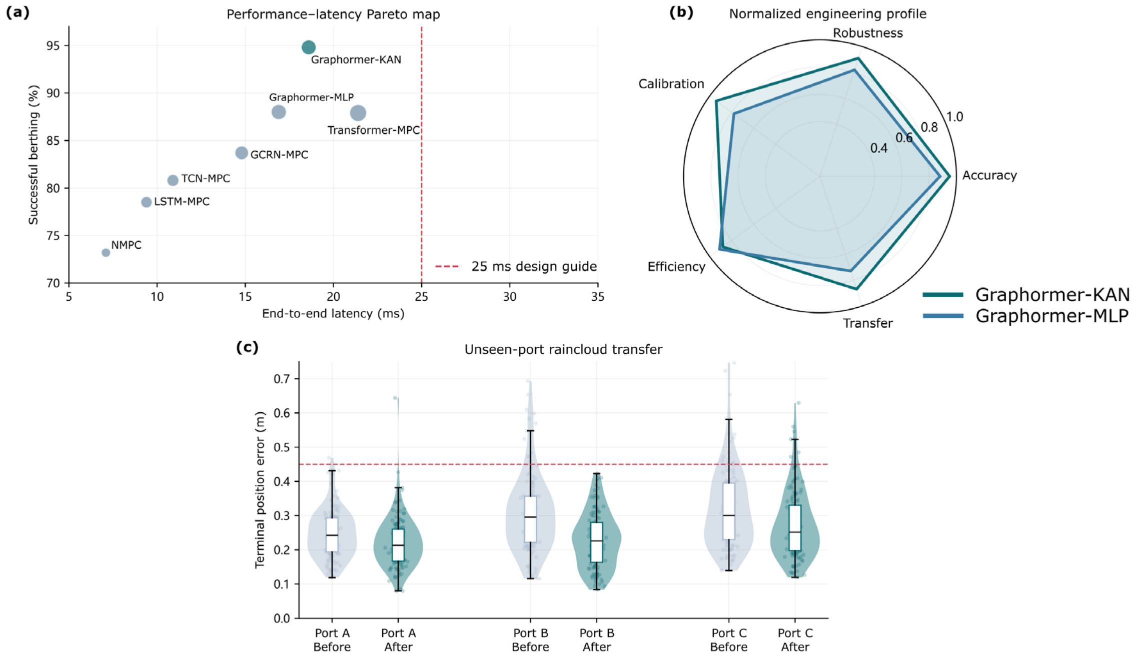

Figure 7. Efficiency and cross-port transfer: (a) Pareto scatter of success, latency, and parameter count; (b) radar profile across five engineering criteria; (c) raincloud terminal-error distributions for three unseen ports before and after light calibration.

Sensitivity analysis varied graph window length, knot count and base margin. Windows shorter than 0.8 s reduced the effectiveness of yaw-rate estimation; windows longer than 1.6 s increased latency without improving results. Eight spline intervals produced the best validation likelihood; sixteen intervals slightly reduced the training error but worsened unseen-port calibration. Margins below 0.25 m increased the contact probability, and margins above 0.55 m extended timeouts in the recessed berth. A safety study shows that a conservative controller may appear safer by avoiding difficult turns [26]. Therefore, the selected configuration will report both collision and completion. The evidence supports an engineering trade-off: only when perception or motion prediction is unreliable does enlarged margins of calibrated uncertainty increase the margin; otherwise, maneuver completion is maintained.

Failure reconstruction offers supporting engineering information. In positioning-bias contacts, the graph is internally consistent but shifted with respect to the berth map; thus, attention cannot correct a common-mode localization error without an external geometric anchor. Radar-Lidar disagreement generates a different signature: dynamic-object covariance increases, but the association module occasionally switches identities and adds discontinuous relative velocity. The residual gate shuts down most of these switches; however, three episodes still enter the safety corridor before the confidence increases. Late obstacle incursion is mainly a planning horizon problem rather than a prediction error. The safety layer will immediately initiate braking, but the remaining lateral authority cannot cope with the simultaneous gust. The baseline contacts are distributed more widely in the nominal and single-fault conditions. Transformer-MPC often focuses on distant high-confidence objects and neglects a partially visible quay corner, and NMPC accumulates unmodeled transverse velocity during low-speed rotation. Logs also show that KAN spline activations stay within their trained coordinate range in all but 0.7% of test calls. Clamp these calls to the validated range and generate a conservative but continuous command.

Resource measurements were repeated over 20,000 online calls after a warm-up period. Median graph assembly requires 3.1 ms, neural inference 9.8 ms, conformal scaling 0.2 ms, and the quadratic-program safety update 5.5 ms. The 95th-percentile end-to-end time is 27.4 ms and the maximum is 63.8 ms. Peak device memory is 612 MB, of which temporal token storage accounts for 38%. Pruning negligible spline coefficients reduces decoder operations by 17% with a 0.3 percentage-point change in success. Half-precision attention is used at deployment, whereas covariance formation and the safety optimizer remain in single precision to avoid loss of positive definiteness. Energy draw rises by 7.6 W above the idle embedded computer during continuous inference, a modest fraction of the vessel hotel load. These measurements indicate that the model does not require a shore-side accelerator. They also expose the optimization solver, rather than attention, as the largest source of latency variation when many boundary constraints become active.

Cross-port behavior was examined qualitatively as well as numerically. The first held-out port contains curved quay geometry, the second uses sparsely spaced cylindrical fenders, and the third includes a moving service lane beside the berth. Before calibration, attention maps still concentrate on hull endpoints and the nearest reachable quay segments, suggesting that relation encodings transfer across map appearance. The largest error occurs at the service lane because obstacle intent is not represented explicitly; the model assigns wide covariance when a craft slows near the approach corridor. Twenty calibration episodes mainly shift the KAN functions associated with transverse load and actuator effectiveness, while relation-bias embeddings change only through frozen normalization. Improvements therefore arise from modest adaptation of vessel–environment response rather than memorization of new geometry. A leave-one-layout-out test inside the original port gives a similar pattern: terminal error increases most for the recessed slip, but calibration recovers approximately half of the loss. This result supports deployment through a limited commissioning maneuver set, provided that the new port supplies an accurate local map and contains no object classes absent from training.

Conclusion

This paper proposes a port-perception-sequence-driven Graphormer-KAN framework for autonomous surface vessel berthing. Organize heterogeneous observations into dynamic graphs, using relation-biased spatial attention and causal temporal aggregation to preserve port-scene interactions, and employ a spline-based KAN decoder to generate motion and actuator distributions. Conformal covariance scaling and an uncertainty-conditioned receding-horizon controller link predictive confidence to executable clearance constraints. The experimental program included prediction, calibration, closed-loop safety, component contribution, computation and transfer under different berth geometries and sensing conditions. Together, these factors provide a reasonable architecture for perception-to-control that distributes the functions of relational structure and local nonlinear decoding separately.

There are still several defects. In this study, a high-fidelity planar simulator and parameterization of a single ship category were used; therefore, unmodeled three-dimensional wave effects, propeller ventilation, and real sensor artifacts may affect field performance. The construction of the graph also requires a stable object association and an available berth-fixed map; extended track swaps or map errors will corrupt the temporal sequence. Distributional decoding assumes that the validation residuals are still reasonably representative for conformal scaling, but this assumption may be violated after sudden hardware failures or severe environmental changes. Finally, the safety layer can only ensure the modelled geometric constraints and cannot guarantee avoidance when facing an adversarial obstacle intention or lost actuation authority.

Future work will link the architecture to hardware-in-the-loop and progressive sea trials, starting with supervised shadow operation and gradually increasing autonomy. Online map correction and identity-robust temporal association can reduce the need for prepared port infrastructure. A richer hydrodynamic token may incorporate water depth, bank suction and learned propulsor health, and set-valued or multimodal prediction can represent ambiguous obstacle behaviour more faithfully than a single Gaussian family. Add formal runtime assurance, fault-detection logic and a verified fallback controller before unattended deployment. Cross-vessel adaptation without calibration maneuvers and privacy-preserving fleet learning will also explore whether the learned spline functions contain transferable maneuvering structure rather than vessel-specific correlations.

Author Contributions

Hamdan Al-Suwaidi contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Ahmed Al-Mansoori contributes to analysis, investigation, data collection, draft preparation.

Rashid Al-Jabri contributes to methodology, visualization. Zayed Al-Tamimi contributes to visualization. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Ziajka-Poznańska, E., & Montewka, J. (2021). Costs and benefits of autonomous shipping—A literature review. Applied Sciences, 11(10), 4553. https://doi.org/10.3390/app11104553

[2] Du, Z., Reppa, V., & Negenborn, R. R. (2020). Cooperative control of autonomous tugs for ship towing. IFAC-PapersOnLine, 53(2), 14470–14475. https://doi.org/10.1016/j.ifacol.2020.12.1448

[3] Chen, C., & Li, Y. (2021). Ship berthing information extraction system using three-dimensional light detection and ranging data. Journal of Marine Science and Engineering, 9(7), 747. https://doi.org/10.3390/jmse9070747

[4] Wróbel, K., Gil, M., & Montewka, J. (2020). Identifying research directions of a remotely-controlled merchant ship by revisiting her system-theoretic safety control structure. Safety Science, 129, 104797. https://doi.org/10.1016/j.ssci.2020.104797

[5] Bitar, G., Martinsen, A. B., Lekkas, A. M., & Breivik, M. (2020). Trajectory planning and control for automatic docking of ASVs with full-scale experiments. IFAC-PapersOnLine, 53(2), 14488–14494. https://doi.org/10.1016/j.ifacol.2020.12.1451

[6] Pedersen, T. A., Glomsrud, J. A., Ruud, E. L., Simonsen, A., Sandrib, J., & Eriksen, B. O. H. (2020). Towards simulation-based verification of autonomous navigation systems. Safety Science, 129, 104799. https://doi.org/10.1016/j.ssci.2020.104799

[7] Blindheim, S., Gros, S., & Johansen, T. A. (2020). Risk-based model predictive control for autonomous ship emergency management. IFAC-PapersOnLine, 53(2), 14524–14531. https://doi.org/10.1016/j.ifacol.2020.12.1456

[8] Zhang, M., Zhang, D., Yao, H., & Zhang, K. (2020). A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration. Safety Science, 130, 104838. https://doi.org/10.1016/j.ssci.2020.104838

[9] Ventikos, N. P., Chmurski, A., & Louzis, K. (2020). A systems-based application for autonomous vessels safety: Hazard identification as a function of increasing autonomy levels. Safety Science, 131, 104919. https://doi.org/10.1016/j.ssci.2020.104919

[10] Park, S., Cap, M., Alonso-Mora, J., Ratti, C., & Rus, D. (2021). Social trajectory planning for urban autonomous surface vessels. IEEE Transactions on Robotics, 37(2), 452–465. https://doi.org/10.1109/TRO.2020.3031250

[11] Iancu, B., Soloviev, V., Zelioli, L., & Lilius, J. (2021). ABOships—An inshore and offshore maritime vessel detection dataset with precise annotations. Remote Sensing, 13(5), 988. https://doi.org/10.3390/rs13050988

[12] Chan, J. P., Norman, R., Pazouki, K., & Golightly, D. (2022). Autonomous maritime operations and the influence of situational awareness within maritime navigation. WMU Journal of Maritime Affairs, 21(2), 121–140. https://doi.org/10.1007/s13437-022-00264-4

[13] Shi, J.-H., & Liu, Z.-J. (2020). Deep learning in unmanned surface vehicles collision-avoidance pattern based on AIS big data with double GRU-RNN. Journal of Marine Science and Engineering, 8(9), 682. https://doi.org/10.3390/jmse8090682

[14] Felski, A., & Zwolak, K. (2020). The ocean-going autonomous ship—Challenges and threats. Journal of Marine Science and Engineering, 8(1), 41. https://doi.org/10.3390/jmse8010041

[15] Chae, C.-J., Kim, M., & Kim, H.-J. (2020). A study on identification of development status of MASS technologies and directions of improvement. Applied Sciences, 10(13), 4564. https://doi.org/10.3390/app10134564

[16] Hu, Y., Zhang, A., Tian, W., Zhang, J., & Hou, Z. (2020). Multi-ship collision avoidance decision-making based on collision risk index. Journal of Marine Science and Engineering, 8(9), 640. https://doi.org/10.3390/jmse8090640

[17] Zhao, C., Thies, P., Lars, J., & Cowles, J. (2021). ROV launch and recovery from an unmanned autonomous surface vessel—Hydrodynamic modelling and system integration. Ocean Engineering, 232, 109019. https://doi.org/10.1016/j.oceaneng.2021.109019

[18] Hinostroza, M., Xu, H., & Guedes Soares, C. (2021). Experimental results of the cooperative operation of autonomous surface vehicles navigating in complex marine environment. Ocean Engineering, 219, 108256. https://doi.org/10.1016/j.oceaneng.2020.108256

[19] Boscaino, V., Odetti, A., Marsala, G., Di Cara, D., Panzavecchia, N., Caccia, M., & Tinè, G. (2021). A fuel cell powered autonomous surface vehicle: The Eco-SWAMP project. International Journal of Hydrogen Energy, 46(39), 20732–20749. https://doi.org/10.1016/j.ijhydene.2021.03.158

[20] Wang, R., Li, D., & Miao, K. (2020). Optimized radial basis function neural network based intelligent control algorithm of unmanned surface vehicles. Journal of Marine Science and Engineering, 8(3), 210. https://doi.org/10.3390/jmse8030210

[21] Wang, S., Zhang, Y., & Li, L. (2020). A collision avoidance decision-making system for autonomous ship based on modified velocity obstacle method. Ocean Engineering, 215, 107910. https://doi.org/10.1016/j.oceaneng.2020.107910

[22] Ye, X., Fang, S., Sun, F., Zhang, C., & Xiang, S. (2022). Meta graph transformer: A novel framework for spatial–temporal traffic prediction. Neurocomputing, 491, 544–563. https://doi.org/10.1016/j.neucom.2021.12.033

[23] Van Molle, P., Verbelen, T., Vankeirsbilck, B., De Vylder, J., Diricx, B., Kimpe, T., Simoens, P., & Dhoedt, B. (2021). Leveraging the Bhattacharyya coefficient for uncertainty quantification in deep neural networks. Neural Computing and Applications, 33(16), 10259–10275. https://doi.org/10.1007/s00521-021-05789-y

[24] Wang, D., Yu, J., Chen, L., Li, X., Jiang, H., Chen, K., Zheng, M., & Luo, X. (2021). A hybrid framework for improving uncertainty quantification in deep learning-based QSAR regression modeling. Journal of Cheminformatics, 13(1), 69. https://doi.org/10.1186/s13321-021-00551-x

[25] Zhang, Z., Li, Y., Song, H., & Dong, H. (2021). Multiple dynamic graph based traffic speed prediction method. Neurocomputing, 461, 109–117. https://doi.org/10.1016/j.neucom.2021.07.052

[26] Li, G., Knoop, V. L., & van Lint, H. (2021). Multistep traffic forecasting by dynamic graph convolution: Interpretations of real-time spatial correlations. Transportation Research Part C: Emerging Technologies, 128, 103185. https://doi.org/10.1016/j.trc.2021.103185