Journal of *Marine Science and Engineering*

MDPI

Article

Intelligent Vessels Localization Based on Adaptive Correlation Information Filter Network in Complex Marine and Port Environments

Lei Yan [1,2,*], Wei Zeng [1], Zhixin Xia [3,4], Bo Meng [2], Junli Ge [3] and Deming Kong [4,5]

[1] Hebei Key Laboratory of Marine Perception Network and Data Processing, Northeastern University, Qinhuangdao 066004, China
[2] Hebei Port Group Co., Ltd., Qinhuangdao 066000, China
[3] Shulian Technologies Limited, Hebei Port Group Co., Ltd., Xiong'an New Area 071701, China
[4] School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China
[5] Hebei Yansoft Information System Co., Ltd., Qinhuangdao 066004, China
* Correspondence: yanlei@qhd.neu.edu.cn or ysuyanl@163.com

Abstract

Accurate and robust localization is essential for intelligent vessels operating in complex marine and port environments. However, single-sensor localization is often affected by limited observation range, environmental occlusion, local interference, and sensor degradation. Although multi-sensor fusion can improve localization reliability, unknown cross-correlated measurement noise arising from shared disturbances, time synchronization errors, communication delays, and inconsistent fusion rates may degrade traditional information-filter-based fusion methods. To address this problem, this paper proposes an Adaptive Correlation Information Filter Network (ACIFNet) for multi-sensor fusion localization of intelligent vessels. ACIFNet preserves the recursive structure of the extended information filter and uses a Transformer-based network to learn adaptive information-domain fusion weights, thereby compensating for unknown inter-sensor correlations without explicitly estimating the full correlation covariance matrix. Experiments on constant-velocity, coordinated-turn (CV), and three-degree-of-freedom vessel motion models, together with a real-world restricted-waterway dataset, demonstrate that ACIFNet achieves higher localization accuracy and stability than Edge Incorporative Fusion (EIF)-inexact fusion, measurement fusion, and KalmanNet. In the CV and three-degree-of-freedom experiments, ACIFNet reduces the mean RMSE by 48.7%, 23.2%, and 26.1%, respectively, compared with KalmanNet. On the real-world dataset, ACIFNet achieves a mean position error of 9.90 m, an RMSE of 11.24 m, and a cross-track error of 8.72 m. These results show that ACIFNet effectively combines the interpretability of information filtering with the adaptive representation capability of neural networks for robust multi-sensor fusion localization under unknown cross-correlated measurement noises.

Keywords: intelligent vessel; multi-sensor fusion; information filter; cross-correlated measurement noise; transformer; state estimation; localization

Check for updates

Academic Editor: Zhiguang Feng

Received: 20 June 2026
Revised: 2 July 2026
Accepted: 3 July 2026
Published: 7 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Intelligent vessels, including assisted-driving tugboats and unmanned surface vehicles (USVs), are increasingly deployed in port waters, restricted waterways, maritime inspection, and autonomous navigation support [1,2]. In these scenarios, accurate and robust

localization is a fundamental prerequisite for safe navigation, collision avoidance, path following, and reliable mission execution. Although GNSS/RTK, inertial measurement units, radar, LiDAR, cameras, and shore-based ranging or bearing sensors can provide useful positioning information, a single sensor is often affected by environmental sensitivity, local interference, signal occlusion, restricted observation range, or temporary degradation [3]. Therefore, multi-sensor fusion has become an important approach for improving the accuracy, continuity, and reliability of intelligent vessel localization.

Filtering-based methods provide an interpretable and recursive framework for state estimation. The Kalman filter and its nonlinear extensions, such as the extended Kalman filter and the unscented Kalman filter, have been widely used in navigation and tracking tasks. In multi-sensor systems, the information filter is particularly attractive because local information contributions can be accumulated in the information domain. However, most conventional information-filter-based fusion methods rely on the assumption that measurement noises from different sensors are mutually independent, or that their cross-correlation structure is known in advance [4]. This assumption may be violated in practical marine and port environments. Sensors installed on the same vessel or monitoring the same water area may be affected by common environmental disturbances, multipath reflections, imperfect time synchronization, communication delay, installation coupling, and inconsistent sensor update rates. These factors can introduce unknown cross-correlated measurement noises. If such correlations are ignored, the fusion center may overestimate the amount of independent information, leading to inappropriate fusion weights, overly confident covariance estimates, and degraded localization performance.

Several methods have been developed to address unknown or uncertain correlations in distributed sensor fusion. Covariance Intersection (CI) provides a conservative solution when the cross-correlation between estimates is unknown by avoiding the use of unknown cross-covariance terms [5]. Generalized Covariance Intersection (GCI) extends this idea to more general probability-density fusion and multi-object fusion scenarios [6]. These methods can maintain consistency under unknown correlations, but they are often conservative because they mainly prevent uncertainty underestimation rather than learning the effective time-varying correlation structure from data. Track-to-track fusion and decentralized or federated Kalman filtering improve modularity by allowing local filters to process measurements independently before fusion-center updating [7]. However, their optimality and consistency still depend on how local estimation-error correlations are modeled, approximated, or bounded. In intelligent-vessel localization systems, these correlations may vary with the waterway environment, vessel motion, sensor status, and communication conditions, making fixed analytical modeling difficult.

Robust filtering methods, including H∞ filtering and related robust Kalman-filter variants, reduce sensitivity to model uncertainty and non-Gaussian disturbances through worst-case or bounded-energy formulations [8]. These methods improve robustness when noise statistics are inaccurate, but they usually do not explicitly exploit the hidden inter-sensor correlation information contained in local information increments. In addition, they are mainly model-based and may require conservative parameter tuning. Learning-based state estimation methods have recently attracted increasing attention. KalmanNet preserves the recursive Kalman-filter structure while using a neural network to learn state-update behavior under nonlinear dynamics and partially known models [9]. Other neural-aided filtering methods further combine recurrent networks, attention mechanisms, or Transformer modules with Kalman filtering to improve adaptability under model mismatch or time-varying noise [10,11]. Nevertheless, most existing neural-aided filtering methods focus on learning the Kalman gain, state correction, or noise covariance within a single filtering

process. They do not directly address the fusion-center problem in which multiple local information contributions are coupled by unknown cross-sensor measurement correlations.

Recent studies on robust navigation and localization further indicate the importance of uncertainty-aware fusion in complex environments. Time-synchronization compensation in GNSS/IMU integrated navigation shows that synchronization errors can affect the consistency of multi-source localization information [12]. Robust localization under complex interference demonstrates that environmental disturbances and unreliable measurements may cause serious localization degradation if they are not properly considered [13]. Cooperative AUV localization based on optimized particle filtering also indicates that uncertainty-aware fusion is important in marine multi-agent sensing scenarios [14]. Environment-adaptive GNSS/INS positioning methods further show that robustness-aware and learning-assisted uncertainty weighting can improve positioning reliability in challenging environments [15]. These studies provide useful motivation for robust localization and integrated navigation. However, they mainly focus on synchronization compensation, interference-resistant localization, cooperative localization, or adaptive GNSS/INS integration, while the information-domain compensation of unknown cross-sensor measurement-noise correlations at the fusion center remains insufficiently explored.

To address this issue, this paper proposes an Adaptive Correlation Information Filter Network (ACIFNet) for multi-sensor fusion localization of intelligent vessels in complex marine and port environments. The key idea is to retain the recursive and interpretable structure of the extended information filter while transforming the influence of unknown cross-correlated measurement noise into adaptive fusion weight learning in the information domain. Instead of explicitly estimating the full cross-correlation covariance matrix, ACIFNet learns the effective contribution of each local sensor estimate to the global information update. Specifically, each local sensor first performs an EKF-based prediction and update using its own measurement. The fusion center then converts the local posterior estimates and covariance matrices into information-domain quantities, including local information contributions, information matrices, and temporal information increments. A Transformer-based neural network is used to model temporal dependencies and inter-sensor coupling from these features. The learned fusion weight matrices are finally embedded into the weighted EIF update to obtain the fused state estimate.

Compared with CI and GCI, ACIFNet is less conservative because it learns adaptive information weights from data rather than only ensuring consistency under unknown correlations. Compared with track-to-track fusion and federated Kalman filtering, ACIFNet does not require an explicit prior model of local estimation-error correlations. Compared with robust and H∞ filtering, ACIFNet focuses on correlation-aware information contribution adjustment rather than only worst-case disturbance attenuation. Compared with KalmanNet and other neural-aided Kalman filters, ACIFNet does not replace the local filtering recursion or directly learn the entire state estimator. Instead, it learns the correlation-dependent weighting effect at the fusion center while preserving the local EKF and global EIF structures.

The main contributions of this paper are summarized as follows.

- A learning-assisted information-filter-based fusion framework is developed for intelligent vessel localization under unknown cross-correlated measurement noises. The proposed method preserves the recursive structure of the extended information filter and introduces adaptive fusion weights to compensate for unknown inter-sensor correlations.
- The influence of unknown cross-sensor noise correlation is formulated as an information-domain fusion weight learning problem. This formulation avoids explicit estimation of the full cross-correlation covariance matrix, which is difficult to

obtain accurately in practical marine environments due to its high dimension and time-varying characteristics.

- A Transformer-based ACIFNet architecture is designed to learn adaptive fusion weight matrices from local information contributions, information matrices, and temporal information increments. The self-attention mechanism enables the model to capture temporal evolution and inter-sensor coupling during the fusion process.
- Extensive experiments are conducted on CV, coordinated-turn, and three-degree-of-freedom vessel motion models, together with a real-world restrict-ed-waterway dataset. Quantitative comparisons, robustness analysis, ablation studies, sensitivity analysis, and computational evaluation are provided to verify the localization accuracy, stability, and practical feasibility of the proposed method.

The remainder of this paper is organized as follows. Section 2 presents the overview of ACIFNet. Section 3 formulates the fusion localization problem. Section 4 introduces the EIF-based fusion model. Section 5 describes the proposed network architecture and training strategy. Section 6 presents the experimental results. Section 7 concludes the paper with limitations and future directions.

2. Algorithm Overview

The proposed ACIFNet is a model-data hybrid state estimation method for multi-sensor fusion localization. As shown in Figure 1, the overall framework is built on the extended information filter. Unlike conventional EIF-based fusion methods that require accurate prior knowledge of inter-sensor noise correlations, ACIFNet treats the unknown correlation effect as an adaptive fusion weight learning problem.

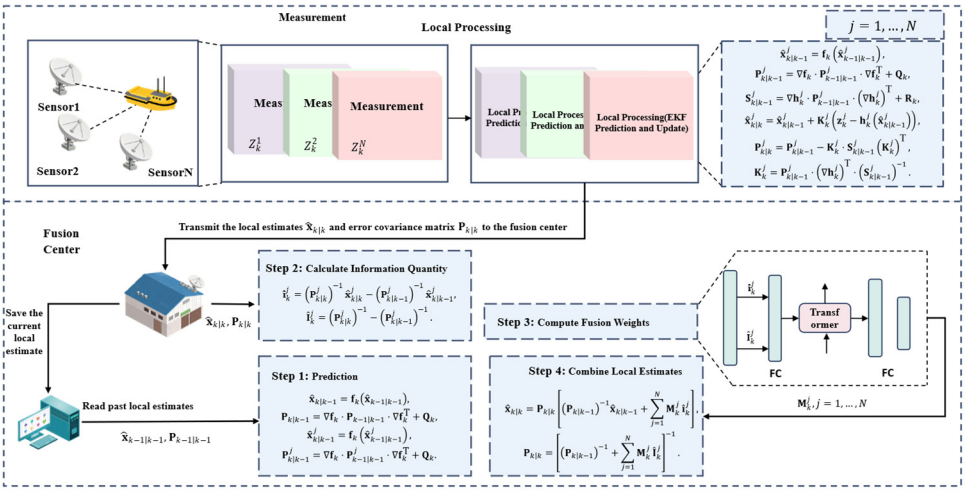


Figure 1. Structure diagram of the ACIFNet system.

The online inference procedure of ACIFNet is summarized in Algorithm 1. The algorithm consists of two stages. In the first stage, each local sensor independently performs an EKF-based prediction and update process using its own measurement. The local posterior state estimate and covariance are then transmitted to the fusion center. In the second stage, the fusion center converts these local estimates into information-domain quantities, including local information contributions and local information matrices. These quantities contain implicit information about the measurement update strength, uncertainty level, and temporal variation of each sensor. A Transformer-based network is then used to learn adaptive fusion weights for each sensor. Finally, the learned weights are embedded into the EIF update equations to obtain the fused state estimate.

In this way, ACIFNet preserves the recursive structure and physical interpretability of information filtering, while using neural networks to compensate for the unknown cross-correlation among multi-sensor measurement noises. This design enables the fusion system to maintain high localization accuracy even when the correlation matrix is unavailable or inaccurate.

Algorithm 1. Online inference procedure of ACIFNet

1: Inputs: The state-space model f_k, h_k^i, $\mathbf{Q}_k$, $\mathbf{R}_k^i$, trained ACIFNet parameter Θ, measurements $\{\mathbf{y}_k^i\}_{i=1}^M$, initial state $\hat{\mathbf{x}}_0$, $\mathbf{P}_0$.
2: Outputs: The fused state estimate $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$.
3: **For** $k = 1 : T$ **do**
4: Compute the prior state estimate and covariance:

$$\hat{\mathbf{x}}_{k|k-1} = f_k(\hat{\mathbf{x}}_{k-1|k-1})$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k$$

5: Convert the prior estimate into information form:

$$\boldsymbol{\Omega}_{k|k-1} = \mathbf{P}_{k|k-1}^{-1}, \boldsymbol{\xi}_{k|k-1} = \boldsymbol{\Omega}_{k|k-1} \hat{\mathbf{x}}_{k|k-1}$$

6: **For** $i = 1 : M$ **do:**
7: Local sensor i performs EKF update using $\mathbf{y}_k^i$, and obtains local estimate $\hat{\mathbf{x}}_{k|k}^i$ and covariance $\mathbf{P}_{k|k}^i$.
8: Convert the local estimate into information form:

$$\boldsymbol{\Omega}_k^i = (\mathbf{P}_{k|k}^i)^{-1}, \boldsymbol{\xi}_k^i = \boldsymbol{\Omega}_k^i \hat{\mathbf{x}}_{k|k}^i$$

9: Calculate the local information increments:

$$\Delta\boldsymbol{\Omega}_k^i = \boldsymbol{\Omega}_k^i - \boldsymbol{\Omega}_{k|k-1}, \Delta\boldsymbol{\xi}_k^i = \boldsymbol{\xi}_k^i - \boldsymbol{\xi}_{k|k-1}$$

10: Construct the ACIFNet input feature:

$$\mathbf{z}_k^i = \left[\Delta\boldsymbol{\xi}_k^i, \Delta\boldsymbol{\Omega}_k^i, \boldsymbol{\Omega}_k^i\right]$$

11: **End for**
12: The fusion center calculates adaptive fusion weights by ACIFNet:

$$\{\mathbf{W}_k^i\}_{i=1}^M = \mathrm{ACIFNet}_\Theta(\{\mathbf{z}_k^i\}_{i=1}^M)$$

13: Update the global information matrix and information state:

$$\boldsymbol{\Omega}_{k|k} = \boldsymbol{\Omega}_{k|k-1} + \sum_{i=1}^M \mathbf{W}_k^i \Delta\boldsymbol{\Omega}_k^i$$

$$\boldsymbol{\xi}_{k|k} = \boldsymbol{\xi}_{k|k-1} + \sum_{i=1}^M \mathbf{W}_k^i \Delta\boldsymbol{\xi}_k^i$$

14: Recover the fused covariance and state estimate:

$$\mathbf{P}_{k|k} = \boldsymbol{\Omega}_{k|k}^{-1}, \hat{\mathbf{x}}_{k|k} = \mathbf{P}_{k|k} \boldsymbol{\xi}_{k|k}$$

15: **End for**
16: **Return** $\hat{\mathbf{x}}_{k|k}$, $\mathbf{P}_{k|k}$.

3. Problem Formulation

Following standard Kalman-filter-based state-space modeling [16,17], consider a nonlinear dynamic system observed by N sensors in discrete time. Here, the sensor i generically represents heterogeneous observation modalities, such as shore-based radar, LiDAR, or AIS, each subject to different noise characteristics. This paper focuses on providing a robust multi-sensor cooperative localization paradigm under realistic complex environments. The system state evolution model and the local measurement model of the i-th sensor are defined as:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}) + \mathbf{w}_k, \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k), \tag{1}$$

$$\mathbf{y}_k^i = h_k^i(\mathbf{x}_k) + \mathbf{v}_k^i, \mathbf{v}_k^i \sim \mathcal{N}\left(0, \mathbf{R}_k^i\right). \tag{2}$$

Here $\mathbf{x}_k$ denotes the system state at time step k, $f_k(\cdot)$ is the nonlinear state transition function, $\mathbf{w}_k$ is the process noise with covariance $\mathbf{Q}_k$, $\mathbf{y}_k^i$ is the measurement from the i-th sensor, $h_k^i(\cdot)$ is the corresponding nonlinear observation function, and $\mathbf{v}_k^i$ is the local measurement noise with covariance $\mathbf{R}_k^i$.

In conventional multi-sensor fusion, measurement noises from different sensors are often assumed to be mutually independent. However, in practical localization systems, different measurement noises may be correlated at the same time step. For $i \neq j$, the cross-covariance can be expressed as:

$$\mathbb{E}\left[\mathbf{v}_k^i \left(\mathbf{v}_k^j\right)^T\right] = \mathbf{R}_k^{ij} \neq 0. \tag{3}$$

The objective of this paper is to estimate the system state $\mathbf{x}_k$ from multi-sensor measurements when such cross-correlations exist, but their exact covariance structure is unavailable to the fusion center.

For measurement fusion, the local sensor measurements can be stacked into an augmented measurement model:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}) + \mathbf{w}_k, \mathbf{x}_k \in \mathbb{R}^m, \tag{4}$$

$$\mathbf{y}_k = h_k(\mathbf{x}_k) + \mathbf{v}_k, \mathbf{y}_k \in \mathbb{R}^{Nn}. \tag{5}$$

The stacked measurement vector, observation function, and measurement noise vector are defined as:

$$\mathbf{y}_k = \left[\left(\mathbf{y}_k^1\right)^T, \left(\mathbf{y}_k^2\right)^T, \ldots, \left(\mathbf{y}_k^N\right)^T\right]^T, \tag{6}$$

$$h_k(\cdot) = \left[\left(h_k^1(\cdot)\right)^T, \left(h_k^2(\cdot)\right)^T, \ldots, \left(h_k^N(\cdot)\right)^T\right]^T, \tag{7}$$

$$\mathbf{v}_k = \left[\left(\mathbf{v}_k^1\right)^T, \left(\mathbf{v}_k^2\right)^T, \ldots, \left(\mathbf{v}_k^N\right)^T\right]^T. \tag{8}$$

The stacked measurement noise covariance is:

$$\mathbf{R}_k = \mathrm{Cov}\left(\left[\left(\mathbf{v}_k^1\right)^T, \left(\mathbf{v}_k^2\right)^T, \ldots, \left(\mathbf{v}_k^N\right)^T\right]^T\right). \tag{9}$$

The diagonal blocks of $\mathbf{R}_k$ correspond to the local measurement noise covariances $\mathbf{R}_k^i$, while the off-diagonal blocks contain the cross-covariances $\mathbf{R}_k^{ij}$. In practice, the diagonal blocks may be approximately known, but the off-diagonal blocks are often difficult to model accurately. This unknown correlation structure is the main source of fusion uncertainty considered in this paper.

4. Model-Based Multi-Sensor Fusion

4.1. Extended Information Filter for Measurement Fusion

The extended information filter is the information-form representation of the extended Kalman filter [18]. Instead of directly using the state estimate and covariance, the information filter represents the posterior distribution through the information state and information matrix. This formulation is particularly suitable for multi-sensor fusion because independent information contributions can be directly added in the information domain.

The EIF-based measurement fusion consists of a prediction and an update step. In the prediction step, the prior state estimate and covariance are computed as:

$$\hat{\mathbf{x}}_{k|k-1} = f_k\left(\hat{\mathbf{x}}_{k-1|k-1}\right), \tag{10}$$

$$\mathbf{P}_{k|k-1} = \nabla f_k \mathbf{P}_{k-1|k-1} \nabla f_k^T + \mathbf{Q}_k, \tag{11}$$

where ∇f_k is the Jacobian matrix of the state transition function.

The innovation obtained from the stacked measurement is:

$$\Delta \mathbf{y}_k = \mathbf{y}_k - h_k\left(\hat{\mathbf{x}}_{k|k-1}\right). \tag{12}$$

The equivalent linearized measurement vector is defined as:

$$\tilde{\mathbf{y}}_k = \Delta \mathbf{y}_k + \nabla h_k \hat{\mathbf{x}}_{k|k-1}, \tag{13}$$

where ∇h_k is the Jacobian matrix of the stacked measurement function.

If the full measurement covariance $\mathbf{R}_k$ is known, the posterior estimate can be obtained by

$$\hat{\mathbf{x}}_{k|k} = \mathbf{P}_{k|k}\left[\mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1} + \nabla h_k^T \mathbf{R}_k^{-1}\tilde{\mathbf{y}}_k\right], \tag{14}$$

$$\mathbf{P}_{k|k}^{-1} = \mathbf{P}_{k|k-1}^{-1} + \nabla h_k^T \mathbf{R}_k^{-1} \nabla h_k. \tag{15}$$

when $\mathbf{R}_k^{-1}$ is exactly known, the above formulation gives the optimal EIF-based measurement fusion. However, in real multi-sensor systems, the full cross-correlation matrix is usually unavailable. This motivates the use of estimation fusion and adaptive learning of the unknown correlation effect.

4.2. EIF-Based Estimation Fusion with Uncorrelated Measurement Noises

When measurement noise from different sensors are mutually independent, multi-sensor estimation fusion can be implemented by local EKF updates and a fusion-center EIF update.

For the i-th sensor, the local EKF prediction is:

$$\hat{\mathbf{x}}_{k|k-1}^i = f_k\left(\hat{\mathbf{x}}_{k-1|k-1}^i\right), \tag{16}$$

$$\mathbf{P}_{k|k-1}^i = \nabla f_k \mathbf{P}_{k-1|k-1}^i \nabla f_k^T + \mathbf{Q}_k, \tag{17}$$

$$\hat{\mathbf{y}}_{k|k-1}^i = h_k^i\left(\hat{\mathbf{x}}_{k|k-1}^i\right), \tag{18}$$

$$\mathbf{S}_{k|k-1}^i = \nabla h_k^i \mathbf{P}_{k|k-1}^i \left(\nabla h_k^i\right)^T + \mathbf{R}_k^i. \tag{19}$$

The local update is:

$$\hat{\mathbf{x}}_{k|k}^i = \hat{\mathbf{x}}_{k|k-1}^i + \mathbf{K}_k^i\left(\mathbf{y}_k^i - \hat{\mathbf{y}}_{k|k-1}^i\right), \tag{20}$$

$$\mathbf{P}_{k|k}^{i} = \mathbf{P}_{k|k-1}^{i} - \mathbf{K}_{k}^{i}\mathbf{S}_{k|k-1}^{i}\left(\mathbf{K}_{k}^{i}\right)^{T}, \tag{21}$$

where the local Kalman gain is:

$$\mathbf{K}_{k}^{i} = \mathbf{P}_{k|k-1}^{i}\left(\nabla h_{k}^{i}\right)^{T}\left(\mathbf{S}_{k\,|\,k-1}^{i}\right)^{-1}. \tag{22}$$

Each local sensor sends its posterior estimate $\hat{\mathbf{x}}_{k|k}^{i}$ and covariance $\mathbf{P}_{k|k}^{i}$ to the fusion center. The fusion center then constructs the local information contribution and local information matrix as:

$$\hat{\mathbf{i}}_{k}^{i} = \left(\mathbf{P}_{k\,|\,k}^{i}\right)^{-1}\hat{\mathbf{x}}_{k|k}^{i} - \left(\mathbf{P}_{k\,|\,k-1}^{i}\right)^{-1}\hat{\mathbf{x}}_{k|k-1}^{i}, \tag{23}$$

$$\hat{\mathbf{I}}_{k}^{i} = \left(\mathbf{P}_{k\,|\,k}^{i}\right)^{-1} - \left(\mathbf{P}_{k\,|\,k-1}^{i}\right)^{-1}. \tag{24}$$

If the measurement noises are uncorrelated, the fusion-center posterior is obtained by directly adding all local information increments:

$$\hat{\mathbf{x}}_{k|k} = \mathbf{P}_{k|k}\left[\mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1} + \sum_{i=1}^{N}\hat{\mathbf{i}}_{k}^{i}\right], \tag{25}$$

$$\mathbf{P}_{k|k}^{-1} = \mathbf{P}_{k|k-1}^{-1} + \sum_{i=1}^{N}\hat{\mathbf{I}}_{k}^{i}. \tag{26}$$

This direct-addition form is simple and efficient, but it is no longer optimal when measurement noises are cross-correlated.

4.3. EIF-Based Estimation Fusion with Cross-Correlated Measurement Noises

When measurement noises are cross-correlated [7,19], direct summation of local information contributions is no longer optimal. If the stacked measurement covariance $\mathbf{R}_k$ is known, the optimal estimation fusion can be expressed by introducing correlation-dependent weighting matrices.

Let

$$\nabla h_{k} = \left[\left(\nabla h_{k}^{1}\right)^{T}, \left(\nabla h_{k}^{2}\right)^{T}, \ldots, \left(\nabla h_{k}^{N}\right)^{T}\right]^{T}, \tag{27}$$

and let $\mathbf{R}_{k}^{-1}(:,i)$ denote the i-th column block of $\mathbf{R}_{k}^{-1}$, and let $\left(\nabla h_{k}^{i}\right)^{*}$ denote the pseudo-inverse of the local measurement Jacobian. Then the posterior under cross-correlated measurement noises can be written as:

$$\hat{\mathbf{x}}_{k|k} = \mathbf{P}_{k|k}\left[\mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1} + \sum_{i=1}^{N}\nabla h_{k}^{T}\mathbf{R}_{k}^{-1}(:,i)\left(\nabla h_{k}^{i}\right)^{*}\hat{\mathbf{i}}_{k}^{i}\right], \tag{28}$$

$$\mathbf{P}_{k|k}^{-1} = \mathbf{P}_{k|k-1}^{-1} + \sum_{i=1}^{N}\nabla h_{k}^{T}\mathbf{R}_{k}^{-1}(:,i)\left(\nabla h_{k}^{i}\right)^{*}\hat{\mathbf{I}}_{k}^{i}. \tag{29}$$

The above formulation is algebraically consistent with EIF-based measurement fusion when the required rank conditions are satisfied. Specifically, when the measurement noises are cross-correlated, and the measurement Jacobians have appropriate rank properties, the influence of unknown correlations can be represented by a set of sensor-dependent weighting matrices. When $\mathbf{R}_k$ is block diagonal, the correlated formulation degenerates into the uncorrelated direct-addition form.

Although the above equations provide an optimal model-based form when the full correlation matrix is known, they require accurate prior knowledge of $\mathbf{R}_k$, ∇h_k, and the un-

derlying state-space model. In practice, especially in intelligent vessels localization under complex marine conditions, such prior knowledge is often inaccurate or unavailable. Therefore, this paper introduces a data-driven mechanism to learn the correlation-dependent fusion weights directly from information-domain features.

5. Adaptive Correlation Information Filter Network

5.1. Fusion Framework

To simplify the information fusion update, the covariance-form state estimate is converted into information form. The information matrix and information state are defined as:

$$\mathbf{Y}_{k|k} = \mathbf{P}_{k|k}^{-1}, \tag{30}$$

$$\hat{\mathbf{y}}_{k|k} = \mathbf{Y}_{k|k}\hat{\mathbf{x}}_{k|k}. \tag{31}$$

The prediction step in information-filter-based estimation fusion is:

$$\hat{\mathbf{x}}_{k|k-1} = f_k\left(\hat{\mathbf{x}}_{k-1|k-1}\right), \tag{32}$$

$$\mathbf{P}_{k|k-1} = \nabla f_k \mathbf{P}_{k-1|k-1} \nabla f_k^T + \mathbf{Q}_k, \tag{33}$$

$$\hat{\mathbf{y}}_{k|k-1} = \mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1}, \tag{34}$$

$$\mathbf{Y}_{k|k-1} = \mathbf{P}_{k|k-1}^{-1}. \tag{35}$$

When measurement noises are uncorrelated, the posterior information state and information matrix are:

$$\hat{\mathbf{y}}_{k|k} = \hat{\mathbf{y}}_{k|k-1} + \sum_{i=1}^{N} \hat{\mathbf{i}}_k^i, \tag{36}$$

$$\mathbf{Y}_{k|k} = \mathbf{Y}_{k|k-1} + \sum_{i=1}^{N} \hat{\mathbf{I}}_k^i. \tag{37}$$

When measurement noises are cross-correlated, the posterior information update becomes:

$$\hat{\mathbf{y}}_{k|k} = \hat{\mathbf{y}}_{k|k-1} + \sum_{i=1}^{N} \mathbf{M}_k^i \hat{\mathbf{i}}_k^i, \tag{38}$$

$$\mathbf{Y}_{k|k} = \mathbf{Y}_{k|k-1} + \sum_{i=1}^{N} \mathbf{M}_k^i \hat{\mathbf{I}}_k^i, \tag{39}$$

where $\mathbf{M}_k^i$ is the correlation-dependent fusion weight matrix for the i-th sensor. In the model-based formulation, it is related to the cross-correlation covariance matrix by:

$$\mathbf{M}_k^i = \nabla h_k^T \mathbf{R}_k^{-1}(:, i)\left(\nabla h_k^i\right)^T. \tag{40}$$

However, the direct computation of $\mathbf{M}_k^i$ requires accurate knowledge of the full cross-correlation matrix $\mathbf{R}_k$, which is difficult to obtain in practice. Therefore, ACIFNet learns $\mathbf{M}_k^i$ from data using information-domain features.

The purpose of ACIFNet is not to estimate the full cross-correlation covariance matrix explicitly. In multi-sensor localization, the full stacked covariance matrix contains both diagonal noise covariances and off-diagonal cross-covariance blocks. Its dimension increases with the number of sensors and measurement channels, and its entries may vary with environmental disturbances, synchronization errors, communication delay, and sensor degradation. Direct online estimation of this matrix is therefore difficult and may be

statistically unreliable under limited samples. From the EIF update perspective, however, the unknown correlation mainly affects the effective information contribution of each local posterior estimate. Therefore, ACIFNet reformulates the correlation compensation problem as learning sensor-dependent information weights. This formulation reduces the learning target from a full covariance structure to its effective role in the information fusion update, while preserving the recursive EIF form.

5.2. Input Features

The input features of ACIFNet are designed according to the information-filter fusion structure. Since the goal of the network is to learn the fusion weight matrix, the input features should contain information related to local information contribution, local information uncertainty, and temporal variation.

For the i-th sensor, three types of features are used:

$$F_1 : \mathbf{i}_k^i = \left(\mathbf{P}_{k\,|\,k}^i\right)^{-1} \hat{\mathbf{x}}_{k|k}^i - \left(\mathbf{P}_{k\,|\,k-1}^i\right)^{-1} \hat{\mathbf{x}}_{k|k-1}^i,$$

$$F_2 : \hat{\mathbf{I}}_k^i = \left(\mathbf{P}_{k\,|\,k}^i\right)^{-1} - \left(\mathbf{P}_{k\,|\,k-1}^i\right)^{-1},$$

$$F_3 : \Delta\hat{\mathbf{i}}_k^i = \hat{\mathbf{i}}_k^i - \hat{\mathbf{i}}_{k-1}^i.$$

Here F_1 represents the measurement-induced information contribution derived from the local state estimate. It changes over time and reflects the effective correction introduced by each sensor. F_2 denotes the local information gain matrix and provides statistical information about estimation confidence. F_3 represents the temporal increment of the local information contribution and helps improve robustness to dynamic changes in sensor reliability and correlation.

The feature vector of the j-th sensor at time k is formed by concatenation:

$$\mathbf{f}_k^j = \left[\hat{\mathbf{i}}_k^j, \hat{\mathbf{I}}_k^j, \Delta\hat{\mathbf{i}}_k^j\right]. \tag{41}$$

For N sensors, the feature set at time k is:

$$\Phi_k = \left\{\mathbf{f}_k^1, \mathbf{f}_k^2, \ldots, \mathbf{f}_k^N\right\}. \tag{42}$$

By stacking the above features along the temporal dimension, a spatio-temporal input sequence is constructed for fusion weight learning.

5.3. Transformer-Based Fusion Weight Learning

In a linear system with fixed prior knowledge, the fusion weight $\mathbf{M}_k^i$ may be time-invariant or weakly time-varying. In nonlinear systems, however, $\mathbf{M}_k^i$ changes over time because the Jacobian matrices and the effective correlation structure vary with the system state. Therefore, the network should capture both temporal dependencies and inter-sensor coupling.

ACIFNet employs a Transformer encoder to learn the adaptive fusion weight matrix. The Transformer is suitable for this task because its self-attention mechanism [20] can directly model the dependency among different sensors and different time steps.

First, each sensor feature vector is mapped to a token representation:

$$\mathbf{z}_k^j = \mathrm{Linear}\left(\mathbf{f}_k^j\right). \tag{43}$$

To preserve temporal and sensor identity information, temporal position encoding and sensor identity encoding are added:

$$\tilde{\mathbf{z}}_k^j = \mathbf{z}_k^j + \mathrm{PE}_t(k) + \mathrm{PE}_s(j). \tag{44}$$

The token sequence is then fed into the Transformer encoder:

$$\mathbf{H} = \mathrm{TransformerEncoder}\left(\left\{\tilde{\mathbf{z}}_t^j\right\}_{t\le k, j\le N}\right). \tag{45}$$

The self-attention operation is defined as

$$\mathrm{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}, \tag{46}$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote the query, key, and value matrices, respectively, and d is the feature dimension. The attention weights describe the importance relationship among different time steps and different sensor information sources. Therefore, the network can implicitly learn the correlation structure among sensors.

After the Transformer encoder obtains the hidden representation $\mathbf{H}_k^j$, a linear mapping layer outputs the fusion weight matrix:

$$\mathbf{M}_k^j = \mathrm{Linear}\left(\mathbf{H}_k^j\right). \tag{47}$$

The learned $\mathbf{M}_k^j$ is then used in the information fusion update.

To improve the numerical stability of the learned information-form update, the output of ACIFNet is constrained before being embedded into the fusion equation. In this implementation, the network does not directly generate an arbitrary unconstrained matrix. Instead, the raw output is transformed into a non-negative diagonal weighting matrix through a softplus mapping. The weighted information update is then symmetrized, and a small regularization term is added to the fused information matrix:

$$\mathbf{Y}_k = \overline{\mathbf{Y}}_k + \sum_{i=1}^{M} \mathbf{W}_{i,k}\Delta\mathbf{Y}_{i,k}\mathbf{W}_{i,k}^T + \varepsilon\mathbf{I} \tag{48}$$

Since each local information increment is positive semidefinite and the weighting operation is implemented as a congruence transformation with non-negative weights, the fused information matrix remains symmetric positive semidefinite. The additional term $\varepsilon\mathbf{I}$ is used only for numerical stabilization and covariance recovery. This constraint prevents the learned fusion weights from producing an invalid information matrix while preserving the adaptive weighting capability of ACIFNet.

5.4. Training Strategy

Assume that the training dataset consists of L trajectories:

$$\mathcal{D} = \left\{\left(\mathbf{X}^l, \mathbf{Z}^{1,l}, \mathbf{Z}^{2,l}, \ldots, \mathbf{Z}^{N,l}\right)\right\}_{l=1}^{L}, \tag{49}$$

where:

$$\mathbf{X}^l = \left[\mathbf{x}_1^l, \mathbf{x}_2^l, \ldots, \mathbf{x}_{T_l}^l\right], \tag{50}$$

$$\mathbf{Z}^{j,l} = \left[\mathbf{z}_1^{j,l}, \mathbf{z}_2^{j,l}, \ldots, \mathbf{z}_{T_l}^{j,l}\right], j = 1, 2, \ldots, N. \tag{51}$$

The trainable parameters of ACIFNet are denoted by Θ. The network is trained in a supervised manner by minimizing the mean squared error between the true state and the fused estimate:

$$\mathcal{L}(\Theta) = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{T_l}\sum_{k=1}^{T_l}\mathbf{x}_k^l - \hat{\mathbf{x}}_{k|k}^l\left(\hat{\mathbf{x}}_{k|k}^{j,l}, \mathbf{P}_{k|k}^{j,l}; \Theta\right)_2^2 + \gamma \parallel \Theta \parallel_2^2, \tag{52}$$

where $\gamma > 0$ is a regularization coefficient.

For mini-batch training with B sampled trajectories, the batch loss is:

$$\mathcal{L}_B(\Theta) = \frac{1}{B}\sum_{l=1}^{B}\frac{1}{T_l}\sum_{k=1}^{T_l}\mathbf{x}_k^l - \hat{\mathbf{x}}_{k|k}^l\left(\hat{\mathbf{x}}_{k|k}^{j,l}, \mathbf{P}_{k|k}^{j,l}; \Theta\right)_2^2 + \gamma \parallel \Theta \parallel_2^2 . \tag{53}$$

The model parameters are updated by gradient descent or its adaptive variants:

$$\Theta \leftarrow \Theta - \alpha \nabla_\Theta \mathcal{L}_B(\Theta), \tag{54}$$

where α is the learning rate.

The training procedure can be summarized as follows. First, a mini-batch of trajectories is randomly sampled from the training dataset. For each trajectory and each time step, every local sensor performs EKF prediction and update using its own measurement. The local estimates and covariance matrices are then transmitted to the fusion center. The fusion center computes the local information contribution and information matrix, and ACIFNet uses these information-domain features to output the adaptive fusion weight matrices. The fusion center then performs weighted information fusion and obtains the fused state estimate. Finally, the mini-batch loss is computed and the network parameters are updated by backpropagation. This process is repeated until the loss converges.

6. Experimental Results

6.1. Experimental Setup

The proposed ACIFNet is evaluated in a sensor network composed of multiple sensor nodes and one fusion center. The measurements of different sensors are assumed to be cross-correlated. For each simulation case, the dataset contains 100 training trajectories, and each training trajectory has 100 time steps. In addition, 20 validation trajectories and 40 testing trajectories are generated. All trajectories have different random initial states.

To ensure a fair comparison, ACIFNet and KalmanNet are trained using the same optimization settings. Both methods adopt the Adam optimizer [21], and Xavier uniform initialization [22] is used to initialize the network weights. The detailed architecture settings and training hyperparameters are listed in Table 1. The following baseline methods are compared.

- IF/EIF-exact: The optimal information fusion method using the true correlation matrix $\mathbf{R}_k$. It is used as the theoretical reference.
- IF/EIF-inexact: The inexact information fusion method that incorrectly assumes mutually independent sensor noises and treats $\mathbf{R}_k$ as block diagonal.
- Measurement fusion: The multi-sensor measurements are stacked into a single large measurement vector, and a centralized IF/EIF update is performed at the fusion center.
- KalmanNet [9]: A representative learning-based measurement fusion method.
- ACIFNet: The proposed method, which learns the adaptive fusion weight matrix $\mathbf{M}_k^i$ from information-domain features and performs weighted information fusion.

In terms of computational complexity, traditional EIF-based fusion mainly involves covariance or information matrix operations, with a typical single-step complexity of

approximately $O(n^3)$. ACIFNet introduces a Transformer module for correlation modeling. During online inference, the additional computational cost mainly comes from the forward propagation of fixed-length sensor features, and the attention module has a complexity of approximately $O(L^2 d)$, where L is the input sequence length and d is the feature dimension. Since the number of sensors and the temporal window length are limited in the experiments, the inference cost remains acceptable.

Table 1. Network architecture and training parameters of ACIFNet and KalmanNet.

Item	ACIFNet	KalmanNet
Main network structure	Transformer encoder + fusion-weight head	GRU-based Kalman gain network
Temporal window length	10	10
Embedding dimension	64	64
Number of encoder/recurrent layers	2	2
Number of attention heads	4	Not applicable
Feed-forward dimension	128	Not applicable
Hidden dimension	64	64
Activation function	ReLU	ReLU
Dropout rate	0.1	0.1
Optimizer	Adam	Adam
Learning rate	1×10^{-3}	1×10^{-3}
Batch size	5	5
Weight decay	1×10^{-5}	1×10^{-5}

A more detailed runtime profile is provided in Section 6.8. The reported total inference time includes the local EKF updates, information-domain feature construction, Transformer forward propagation, and the weighted EIF update at the fusion center.

6.2. Constant-Velocity Model

The constant-velocity model [23] is first used to evaluate the fusion performance of ACIFNet under a linear state evolution model. This model describes the motion of a USV moving smoothly on the horizontal plane over a short time interval. The target state is defined as

$$\mathbf{x}_k = [x, \dot{x}, y, \dot{y}]^T, \tag{55}$$

and the measurement is:

$$\mathbf{z}_k = [x, y]^T. \tag{56}$$

The state transition matrix is:

$$\mathbf{F}_k = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{57}$$

The process noise covariance is configured as:

$$\mathbf{Q}_k = q^2 \begin{bmatrix} T^4/3 & T^3/2 & 0 & 0 \\ T^3/2 & T^2 & 0 & 0 \\ 0 & 0 & T^4/3 & T^3/2 \\ 0 & 0 & T^3/2 & T^2 \end{bmatrix}. \tag{58}$$

Two sensors are used for position measurement, and the observation matrix is:

$$\mathbf{H}_k^j = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, j = 1, 2. \tag{59}$$

The sampling period is $T = 1$, and the process noise parameter is $q = 1$. The cross-correlated measurement noise covariance is set as:

$$\mathbf{R}_k = r^2 \begin{bmatrix} 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0.5 & 0 & 2 & 0 \\ 0 & 0.5 & 0 & 2 \end{bmatrix}. \tag{60}$$

The inaccurate filter parameters are set as $q = 5$ and

$$\mathbf{R}_k^1 = \mathbf{R}_k^2 = \mathrm{diag}(10, 10), \tag{61}$$

which means that the fusion center ignores the true cross-correlation structure.

As shown in Figure 2, after 500 training epochs, the training loss of ACIFNet converges faster and to a lower value than KalmanNet. This indicates that ACIFNet is easier to train under the same data scale. The reason is that KalmanNet directly learns the time-varying Kalman gain, which is complex and highly dynamic, whereas ACIFNet learns the information-domain fusion weight. In the linear case, this weight is time-invariant or weakly time-varying, which reduces the difficulty of training.

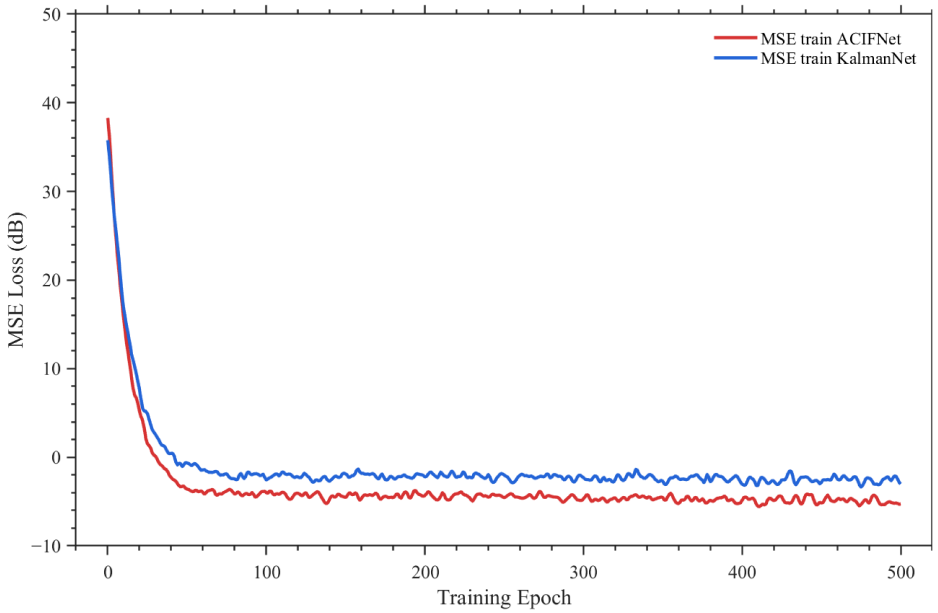


Figure 2. Training loss of ACIFNet and KalmanNet.

In Figure 3, the testing results show that ACIFNet outperforms all baseline methods except the IF-exact reference. Since the system is linear, ACIFNet can effectively extract

the hidden correlation information from training data and learn the corresponding fusion weight. Even when the filter prior parameters and measurement noise correlation structure are inaccurate, ACIFNet compensates for the unknown correlation effect through data-driven weight learning. Its performance approaches the ideal IF-exact baseline and is more stable than KalmanNet, measurement fusion, and IF-inexact.

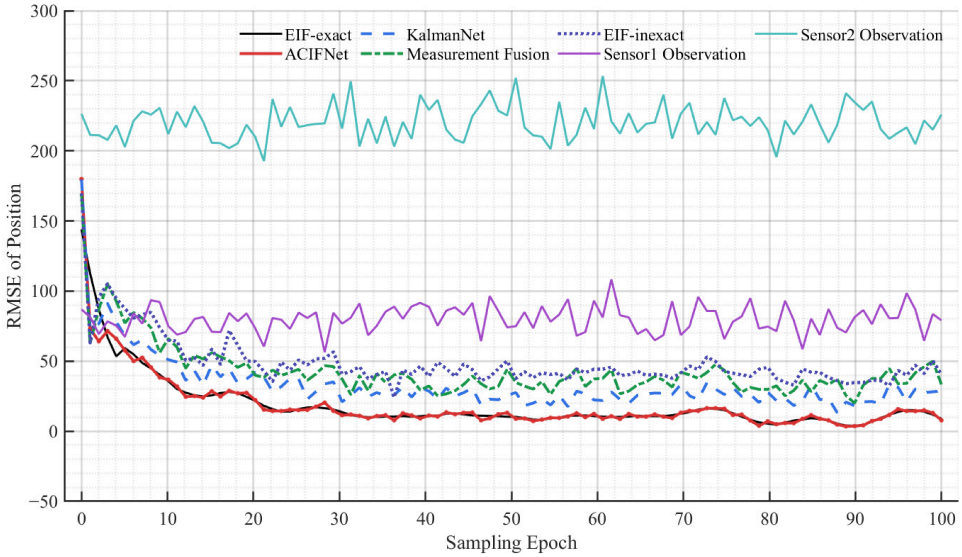


Figure 3. Estimated position RMSE on the test set.

As shown in Table 2, ACIFNet achieves a mean RMSE of 14.88 in the CV model, which is lower than those of EIF-inexact, measurement fusion, and KalmanNet, and approaches the EIF-exact reference. Compared with KalmanNet, ACIFNet reduces the mean RMSE from 29.02 to 14.88, corresponding to an improvement of approximately 48.7%. This result indicates that learning information-domain fusion weights is effective for compensating unknown cross-correlations in the linear CV case.

Table 2. Quantitative results for the CV model.

Method	Mean Error	Std.	Max Error
Sensor 1 observation	80.12	8.94	108.21
Sensor 2 observation	220.90	13.47	254.36
EIF-exact	10.46	3.82	19.15
EIF-inexact	43.67	6.91	45.63
Measurement fusion	38.27	7.85	49.63
KalmanNet	29.02	8.76	41.55
ACIFNet	14.88	4.05	20.74

6.3. Coordinated-Turn Model

To further evaluate the performance under nonlinear dynamics and nonlinear measurements, the coordinated-turn model is considered. This model extends the horizontal motion state by including the turn rate and can better describe maneuvering motion during turning.

The state vector is:

$$\mathbf{x}_k = \left[x, \dot{x}, y, \dot{y}, \Omega\right]^T, \tag{62}$$

where x and y are the target positions, $\dot{x}$ and $\dot{y}$ are the velocity components, and Ω is the turn rate.

The state transition model is:

$$\mathbf{x}_k = \begin{bmatrix} 1 & \frac{\sin \Omega_{k-1} T}{\Omega_{k-1}} & 0 & -\frac{1-\cos \Omega_{k-1} T}{\Omega_{k-1}} & 0 \\ 0 & \cos \Omega_{k-1} T & 0 & -\sin \Omega_{k-1} T & 0 \\ 0 & \frac{1-\cos \Omega_{k-1} T}{\Omega_{k-1}} & 1 & \frac{\sin \Omega_{k-1} T}{\Omega_{k-1}} & 0 \\ 0 & \sin \Omega_{k-1} T & 0 & \cos \Omega_{k-1} T & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{w}_k. \tag{63}$$

The sampling period is $T = 0.2$. The process noise covariance is:

$$\mathbf{Q}_k = \begin{bmatrix} q_x^2 T^3/3 & q_x^2 T^2/2 & 0 & 0 & 0 \\ q_x^2 T^2/2 & q_x^2 T & 0 & 0 & 0 \\ 0 & 0 & q_y^2 T^3/3 & q_y^2 T^2/2 & 0 \\ 0 & 0 & q_y^2 T^2/2 & q_y^2 T & 0 \\ 0 & 0 & 0 & 0 & q_\Omega^2 T \end{bmatrix}, \tag{64}$$

where $q_x = q_y = 5$ and $q_\Omega = 10^{-3}$.

Each sensor measures range and bearing:

$$\mathbf{z}_k^j = \begin{bmatrix} d_k^j \\ A_k^j \end{bmatrix} = \begin{bmatrix} \sqrt{x_k^2 + y_k^2} \\ \tan^{-1}(y_k / x_k) \end{bmatrix} + \mathbf{v}_k^j, j = 1, 2. \tag{65}$$

The correlated measurement noise is generated by adding a shared disturbance:

$$\tilde{\mathbf{v}}_k^j = \mathbf{v}_k^j + \beta_j \mathbf{v}_k^0, \tag{66}$$

where β_j is the scaling coefficient and $\mathbf{v}_k^0$ is the common disturbance. The covariance block of the stacked measurement noise is:

$$\mathbf{R}_k^{ij} = \mathrm{Cov}\left(\tilde{\mathbf{v}}_k^i, \tilde{\mathbf{v}}_k^j\right) = \tilde{\mathbf{R}}_k^{ij} + \beta_i \beta_j \mathbf{R}_k^0. \tag{67}$$

The local independent noise covariance is:

$$\tilde{\mathbf{R}}_k^{ij} = \begin{cases} \begin{bmatrix} r_d^2 & 0 \\ 0 & r_A^2 \end{bmatrix}, & i = j, \\ 0, & i \neq j, \end{cases} \tag{68}$$

and the common disturbance covariance is:

$$\mathbf{R}_k^0 = \begin{bmatrix} r_d^2 & 0 \\ 0 & r_A^2 \end{bmatrix}. \tag{69}$$

The scaling coefficient is $\beta_j = 0.5$, and the measurement noise parameters are $r_d = 10$ and $r_A = 10^{-4}$. For inaccurate filtering, the parameters are set as $q_x = q_y = 10$, $q_\Omega = 2 \times 10^{-3}$, and:

$$\mathbf{R}_k^1 = \mathbf{R}_k^2 = \mathrm{diag}\left(r_d^2, r_A^2\right). \tag{70}$$

When both the state transition model and observation model are nonlinear, the performance differences among different methods become more significant. Under the CT model (in Figure 4), ACIFNet achieves the lowest RMSE and the smallest fluctuation. This demonstrates that ACIFNet does not merely rely on exact model parameters, but can adaptively learn the fusion weights in the information-filter update from multi-sensor

observations. Therefore, it can reduce the influence of nonlinear approximation errors and unknown correlations.

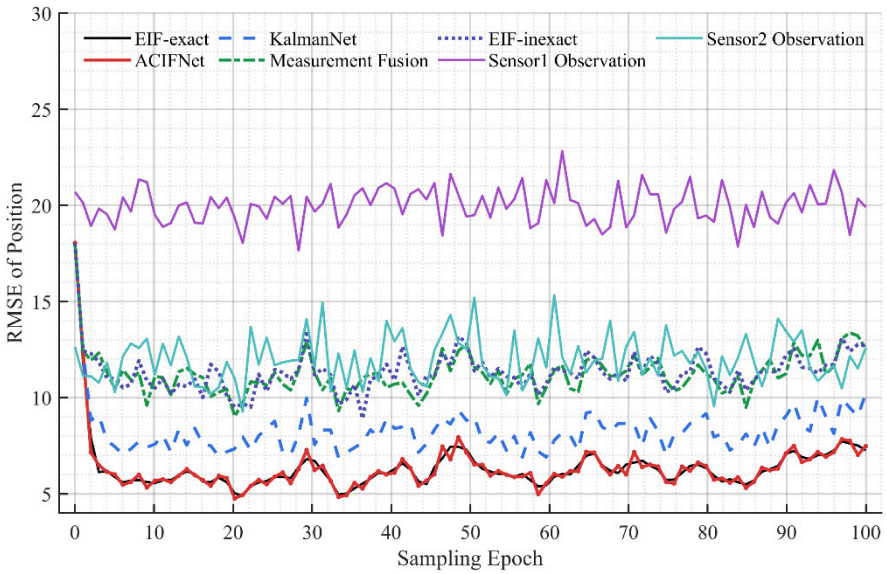


Figure 4. Estimated position RMSE on the test set.

KalmanNet improves estimation accuracy to some extent, but its errors and fluctuations remain higher than those of ACIFNet. Measurement fusion performs close to the raw observation level, indicating that it does not sufficiently exploit the hidden correlation information among multiple sensors. EIF-inexact performs the worst because it ignores the cross-correlated noise structure. These results show that ACIFNet has clear advantages in multi-sensor fusion accuracy and robustness under nonlinear motion and measurement conditions.

Table 3 shows that ACIFNet reduces the mean position RMSE from 8.05 to 6.18 compared with KalmanNet, corresponding to a 23.2% improvement. The standard deviation is also lower, indicating a more stable estimation under nonlinear coordinated-turn motion and range-bearing observations.

Table 3. Quantitative results for the CT model.

Method	Mean Error	Std.	Max Error
Sensor 1 observation	20.58	0.82	22.93
Sensor 2 observation	12.99	1.12	15.76
EIF-exact	5.95	0.63	7.41
EIF-inexact	11.52	0.96	13.28
Measurement fusion	11.38	1.02	13.42
KalmanNet	8.05	0.87	9.82
ACIFNet	6.18	0.74	8.03

6.4. Three-Degree-of-Freedom Motion Model

To further verify the fusion performance of the proposed method under complex nonlinear ship dynamics, a three-degree-of-freedom USV motion model [24] is considered. This model describes planar USV motion in surge, sway, and yaw and is more consistent with practical marine vehicle motion.

The state transition is given by:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \boldsymbol{\tau}_{k-1}) + \mathbf{w}_k, \tag{71}$$

With:

$$f(\mathbf{x}_{k-1}, \boldsymbol{\tau}_{k-1}) = \begin{bmatrix} x_{k-1} + (u_{k-1}\cos\theta_{k-1} - v_{k-1}\sin\theta_{k-1})T \\ y_{k-1} + (u_{k-1}\sin\theta_{k-1} + v_{k-1}\cos\theta_{k-1})T \\ \theta_{k-1} + r_{k-1}T \\ u_{k-1} + \tau_{u,k-1}T \\ v_{k-1} + \tau_{v,k-1}T \\ r_{k-1} + \tau_{r,k-1}T \end{bmatrix}. \tag{72}$$

The process noise covariance is:

$$\mathbf{Q}_k = \mathrm{diag}\left(\sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_u^2, \sigma_v^2, \sigma_r^2\right), \tag{73}$$

where:

$$\mathbf{Q}_k = \mathrm{diag}\left(0.05^2, 0.05^2, 0.01^2, 0.10^2, 0.05^2, 0.02^2\right).$$

Only the position components are observed:

$$\mathbf{y}_k^j = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k^j, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, j = 1, 2. \tag{74}$$

Two sensors with different accuracies and cross-correlated noises are considered. The stacked covariance matrix is configured as:

$$\mathbf{R}_k = \begin{bmatrix} 4 & 0 & 2.828 & 0 \\ 0 & 4 & 0 & 2.828 \\ 2.828 & 0 & 8 & 0 \\ 0 & 2.828 & 0 & 8 \end{bmatrix}. \tag{75}$$

The inaccurate filter parameters are set as:

$$\mathbf{R}_k^1 = \mathbf{R}_k^2 = \mathrm{diag}(4, 4, 4, 4). \tag{76}$$

Under the 3-DOF model, all methods converge rapidly in the initial stage. However, as the estimation process continues, the differences among methods become evident. As shown in Figure 5, ACIFNet maintains the lowest error level throughout the sampling interval, and its error curve is smoother than those of the baseline methods. This indicates that ACIFNet has strong adaptability to complex nonlinear USV dynamics.

KalmanNet and measurement fusion outperform raw single-sensor observations, but their errors are still significantly higher than those of ACIFNet. In particular, measurement fusion shows limited improvement under the 3-DOF model, indicating insufficient use of multi-source correlation information. EIF-inexact suffers the most obvious degradation, further demonstrating that traditional filtering methods are highly dependent on accurate model and noise prior information.

As shown in Table 4, undernder the 3-DOF vessel model, ACIFNet achieves a mean RMSE of 7.06, which is 26.1% lower than KalmanNet and approximately 41.8% lower than EIF-inexact. This result indicates that the proposed adaptive information weighting remains effective under more complex nonlinear vessel dynamics.

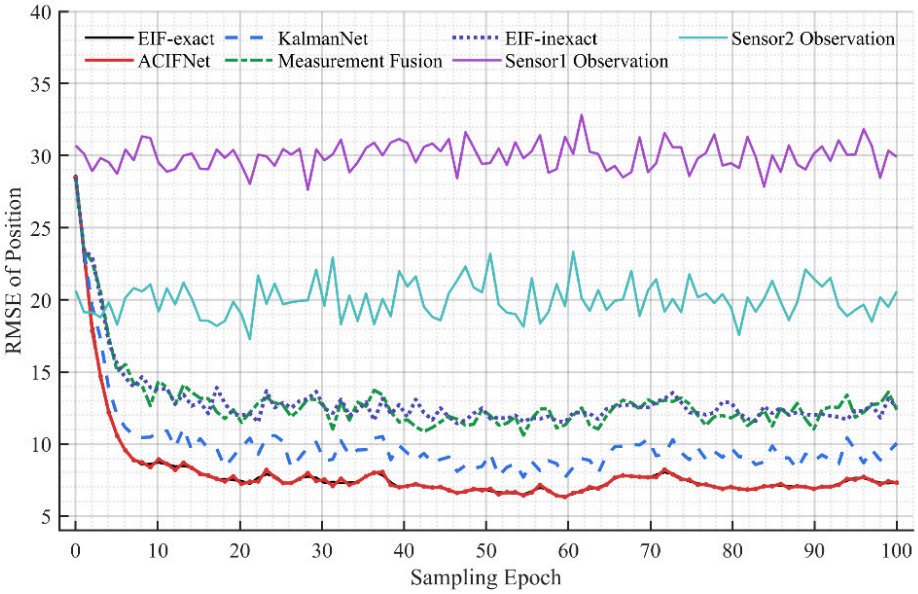


Figure 5. Estimated position RMSE on the test set.

Table 4. Quantitative results for the 3-DOF vessel model.

Method	Mean Error	Std.	Max Error
Sensor 1 observation	30.01	0.89	33.02
Sensor 2 observation	19.76	1.14	23.18
EIF-exact	6.52	0.55	7.84
EIF-inexact	12.14	0.82	13.91
Measurement fusion	12.36	0.91	14.22
KalmanNet	9.55	0.78	10.96
ACIFNet	7.06	0.64	8.42

6.5. Real-World Dataset Experiment

To further evaluate the effectiveness and practical applicability of ACIFNet in a real complex environment, the Pohang Canal Dataset [25] is used. Compared with simulation data, real-world USV localization is affected by more complicated factors, including waterway constraints, GNSS disturbances, environmental reflection, motion uncertainty, and sensor noise variation. Therefore, experiments on real data provide a more comprehensive evaluation of localization accuracy, fusion capability, and robustness.

The state transition model is:

$$f(\mathbf{x}_k) = \begin{bmatrix} x_k + v_k \cos\psi_k \Delta t \\ y_k + v_k \sin\psi_k \Delta t \\ v_k \\ \psi_k \end{bmatrix}, \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}). \tag{77}$$

The true process noise covariance is set as

$$\mathbf{Q}^{true} = \mathrm{diag}\left(\sigma_x^2, \sigma_y^2, \sigma_v^2, \sigma_\psi^2\right), \tag{78}$$

where

$$\mathbf{Q}^{true} = \mathrm{diag}\left(0.01^2, 0.01^2, 0.10^2, (1^\circ)^2\right). \tag{79}$$

GPS-like position and heading observations are used

$$\mathbf{Z}_k^{(j)} = h^{(j)}(\mathbf{x}_k) + \mathbf{v}_k^{(j)} = \begin{bmatrix} x_k^{(j)} \\ y_k^{(j)} \\ \psi_k^{(j)} \end{bmatrix} + \mathbf{v}_k^{(j)}, \mathbf{v}_k^{(j)} \sim \mathcal{N}\left(0, \mathbf{R}^{(j)}\right). \tag{80}$$

To evaluate fusion under cross-correlated disturbances, two observation channels with different independent noise levels and a shared disturbance are considered:

$$\mathbf{Z}_k^{(j)} = \begin{bmatrix} x_k \\ y_k \\ \psi_k \end{bmatrix} + \tilde{\mathbf{v}}_k^{(j)} + \beta_j \mathbf{v}_k^{(0)}. \tag{81}$$

Here $\tilde{\mathbf{v}}_k^{(j)}$ is the independent sensor noise and $\mathbf{v}_k^{(0)}$ is the common disturbance. The independent noise covariance is:

$$\tilde{\mathbf{R}}^{(j)} = \mathrm{diag}\left(\sigma_{p,j}^2, \sigma_{p,j}^2, \sigma_{\psi,j}^2\right). \tag{82}$$

For sensor 1, $\sigma_{p,1} = 0.5\,\mathrm{m}$ and $\sigma_{\psi,1} = 1^\circ$. For sensor 2, $\sigma_{p,2} = 1\,\mathrm{m}$ and $\sigma_{\psi,2} = 2^\circ$. The common disturbance covariance is:

$$\mathbf{R}^{(0)} = \mathrm{diag}\left(1.0^2, 1.0^2, (2^\circ)^2\right). \tag{83}$$

The inaccurate filter parameters are set as:

$$\mathbf{R}^{(j)} = \mathrm{diag}\left(\sigma_{p,j}^2, \sigma_{p,j}^2, \sigma_{\psi,j}^2\right), \tag{84}$$

which means that the filter does not consider cross-sensor noise correlation.

In Figure 6, the experimental results on the real-world dataset show that all methods converge rapidly at the initial stage, but the performance differences become clear in the stable stage. ACIFNet maintains the lowest error level during the whole sampling process, and its fluctuation is smaller than those of the baseline methods. This indicates that ACIFNet achieves higher localization accuracy and better stability in real restricted waterway environments. KalmanNet and measurement fusion can reduce the error of single-sensor observations, but their overall performance is still inferior to ACIFNet. EIF-inexact and measurement fusion show larger errors than the learning-based fusion methods, indicating that inaccurate or insufficient correlation modeling can degrade real-world fusion performance.

In Figure 7, the trajectory estimation results further show that the trajectory estimated by ACIFNet is the closest to the reference trajectory. In turning segments and high-curvature regions, ACIFNet still provides stable and accurate localization. In contrast, KalmanNet and measurement fusion exhibit more pronounced local deviations, indicating limited adaptability to complex real-world motion and observation disturbances. Overall, the real-data experiment verifies that ACIFNet can more effectively fuse multi-source observation information and maintain high localization accuracy and robustness in complex environments.

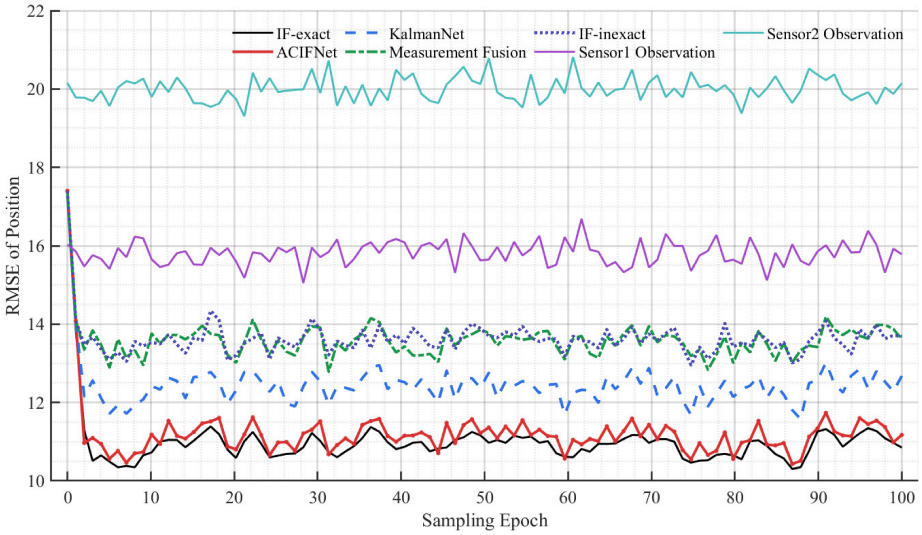


Figure 6. Estimated position RMSE on the test set.

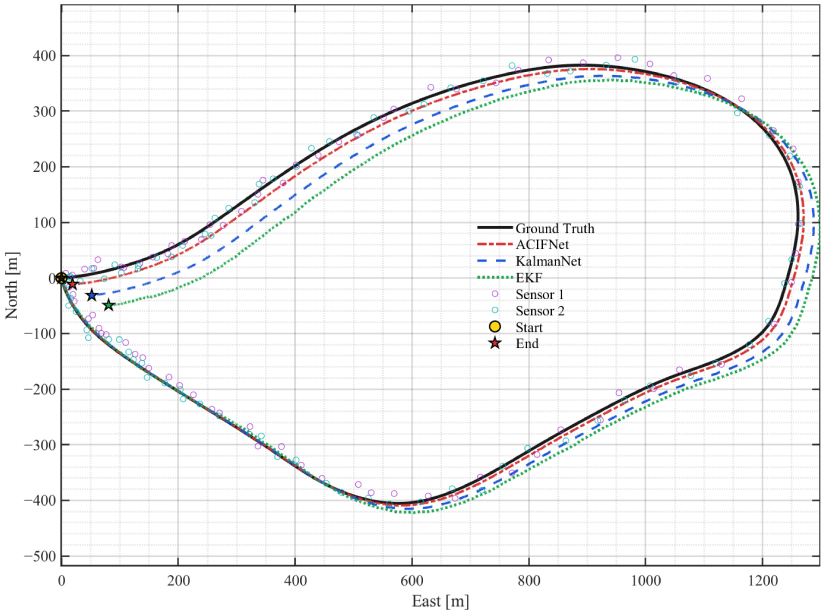


Figure 7. Test trajectory on the Pohang Canal Dataset.

As reported in Table 5, ACIFNet achieves a mean position error of 9.90 m and an RMSE of 11.24 m on the real-world dataset, which are substantially lower than those of KalmanNet, measurement fusion, and IF-inexact. The cross-track error is also reduced to 8.72 m, confirming that ACIFNet provides better lateral trajectory consistency in restricted-waterway navigation.

Table 5. Quantitative results on the real-world dataset.

Method	Mean Position Error/m	RMSE/m	Max Position Error/m	Heading RMSE/deg	Along-Track Error/m	Cross-Track Error/m
Sensor 1 observation	14.14	16.15	16.69	6.82	6.20	12.35
Sensor 2 observation	17.81	20.14	20.80	9.46	7.85	15.65

Table 5. *Cont.*

Method	Mean Position Error/m	RMSE/m	Max Position Error/m	Heading RMSE/deg	Along-Track Error/m	Cross-Track Error/m
EIF-exact	9.69	11.02	11.70	3.21	3.92	8.48
EIF-inexact	12.10	13.71	14.20	5.37	5.12	10.55
Measurement fusion	12.32	13.94	14.35	5.62	5.26	10.78
KalmanNet	11.63	13.18	13.66	4.85	4.82	10.12
ACIFNet	9.90	11.24	11.95	3.48	4.05	8.72

6.6. Robustness Evaluation

To further evaluate the robustness of ACIFNet, three practical sensing uncertainty scenarios are considered: incomplete measurements, sudden environmental disturbances, and sensor degradation. The CT model is used as the representative nonlinear test case.

For incomplete measurements, each sensor measurement is randomly dropped with probability p_{drop}. When the measurement of sensor i is unavailable, its information increment is set to zero:

$$\Delta \boldsymbol{i}_{i,k} = 0, \Delta \mathbf{I}_{i,k} = 0. \tag{85}$$

The dropout probability is set to:

$$p_{\text{drop}} \in \{5\%, 10\%, 30\%\}.$$

For sudden environmental disturbances, a short-term common disturbance is added to the measurements:

$$\mathbf{y}_{i,k}^{d} = \mathbf{y}_{i,k} + \gamma \mathbf{d}_k, k \in [40, 50], \tag{86}$$

where $\gamma \in \{2, 4, 6\}$ is the disturbance intensity.

For sensor degradation, the measurement noise covariance of one sensor is enlarged as

$$\mathbf{R}_i^{d} = \lambda \mathbf{R}_i \tag{87}$$

where $\lambda \in \{2, 4, 8\}$ denotes the degradation factor.

As shown in Table 6, all methods show increased RMSE as the uncertainty level becomes higher. However, ACIFNet consistently achieves the lowest error among the practical fusion methods. This indicates that the proposed adaptive information-weighting mechanism can reduce the influence of missing measurements, short-term shared disturbances, and degraded sensor observations.

Table 6. Robustness comparison under different sensing uncertainty scenarios.

Scenario	Severity	EIF-Inexact	Measurement Fusion	KalmanNet	ACIFNet
Measurement dropout	5%	12.41	12.21	8.82	6.64
Measurement dropout	10%	13.66	13.58	9.91	7.39
Measurement dropout	30%	18.74	18.89	13.90	11.02
Sudden disturbance	$\gamma = 2$	12.05	12.30	8.64	6.73
Sudden disturbance	$\gamma = 4$	14.36	14.65	10.42	8.09
Sudden disturbance	$\gamma = 6$	17.25	17.74	12.3	10.12

Table 6. *Cont.*

Scenario	Severity	EIF-Inexact	Measurement Fusion	KalmanNet	ACIFNet
Sensor degradation	$\lambda = 2$	12.28	12.40	8.55	6.58
Sensor degradation	$\lambda = 4$	14.19	14.77	9.83	7.62
Sensor degradation	$\lambda = 8$	17.66	18.33	12.42	9.68

When the uncertainty becomes severe, the improvement of ACIFNet decreases because the available reliable measurement information is reduced. Therefore, ACIFNet improves robustness under moderate sensing uncertainties, but its performance still depends on the availability of informative sensor measurements.

6.7. Ablation Study

To analyze the contribution of different components in ACIFNet, an ablation study is conducted on the CV, CT, 3-DOF, and real-world experiments. The full ACIFNet is compared with four simplified variants.

First, the Transformer encoder is replaced by a multilayer perceptron, denoted as w/o Transformer, to evaluate the effect of self-attention on temporal and inter-sensor dependency modeling. Second, the temporal increment features are removed, denoted as w/o temporal increment, to verify the contribution of dynamic information variation. Third, the local information matrix is removed from the network input, denoted as w/o information matrix, to examine the importance of uncertainty-related information. Fourth, the adaptive fusion weights are replaced by direct information addition, denoted as w/o adaptive weight, to evaluate whether the learned fusion weights are necessary under unknown cross-correlated measurement noises.

As shown in Table 7, removing any major component leads to performance degradation. The largest degradation occurs when the adaptive fusion weights are removed. In this case, the method degenerates into direct EIF-based information addition and cannot compensate for unknown cross-correlated measurement noises. This confirms that adaptive information weighting is the core component of ACIFNet.

Table 7. Ablation study of ACIFNet under different experimental scenarios. Note: "w/o" denotes "without".

Method Variant	CV RMSE	CT RMSE	3-DOF RMSE	Real-World RMSE/m
Full ACIFNet	14.88	6.18	7.06	11.24
w/o Transformer	22.35	7.42	8.31	12.07
w/o temporal increment	20.61	7.16	8.05	11.90
w/o information matrix	24.84	7.83	8.74	12.34
w/o adaptive weight	43.67	11.52	12.14	13.71

Replacing the Transformer with an MLP also increases the estimation error, especially in the CT, 3-DOF, and real-world experiments. This indicates that self-attention is useful for capturing temporal dependencies and inter-sensor coupling. Removing the temporal increment features also degrades the performance, showing that the variation of information contribution over time helps the network identify changes in sensor reliability and correlation. In addition, removing the information matrix causes clear performance loss because the network no longer receives sufficient uncertainty-related information.

Overall, the ablation results demonstrate that the performance improvement of ACIFNet mainly comes from the adaptive information-weighting mechanism, while the Transformer encoder, temporal increment features, and information matrix features further improve the accuracy and stability of the fusion process.

From the relative degradation perspective, removing the adaptive information weight produces the largest increase in error, with the CV RMSE increasing from 14.88 to 43.67. Removing the information matrix causes the second largest degradation, because the network loses direct uncertainty-related input. Replacing the Transformer with an MLP and removing temporal increments also degrade performance, showing that the gain of ACIFNet is not caused by a single module alone, but mainly by the combination of information-domain weighting, uncertainty features, and temporal inter-sensor dependency modeling.

6.8. Runtime Profile and Sensitivity Analysis

To further evaluate practical feasibility and robustness to correlation-related settings, additional runtime profiling and sensitivity analyses are conducted on the CT model, which contains both nonlinear dynamics and nonlinear range-bearing measurements. All runtime results are measured on the same experimental platform used for the main experiments.

Table 8 reports the average single-step runtime of the main online inference components. The Transformer module accounts for the largest part of the additional cost, but the total average inference time remains 5.60 ms per step. This indicates that the proposed method is suitable for typical multi-sensor localization tasks with moderate sampling rates. For larger sensor networks or embedded platforms, model compression and lightweight attention mechanisms may further reduce the computation burden.

Table 8. Runtime profile of ACIFNet online inference.

Component	Main Operation	Average Time/ms	Percentage
Local EKF inference	Prediction and update of all local sensors	1.18 ± 0.12	21.1%
Feature extraction	Information contribution, matrix, and temporal increment	0.74 ± 0.08	13.2%
Transformer inference	Token embedding, self-attention, and weight head	2.61 ± 0.20	46.6%
Fusion-center update	Weighted EIF update and state recovery	0.83 ± 0.09	14.8%
Other overhead	Data formatting and normalization	0.24 ± 0.03	4.3%
Total	Complete ACIFNet online inference	5.60 ± 0.34	100%

Table 9 reports the sensitivity of the proposed method to the cross-correlation coefficient on the CT model. As the cross-correlation between sensor noises becomes stronger, all methods suffer larger RMSE because the amount of effectively independent information decreases. The error growth of EIF-inexact and measurement fusion is more obvious because they treat correlated information as independent. ACIFNet maintains the lowest RMSE across all tested correlation levels, although its advantage decreases when the correlation becomes extremely strong.

Table 10 evaluates the influence of the common disturbance strength on the CT model. As the common disturbance becomes stronger, the shared perturbation affecting different sensors is amplified, which leads to more severe cross-correlated measurement noise and larger estimation errors for all methods. ACIFNet remains more robust than the compared

practical fusion methods because the learned information weights can adaptively reduce the impact of correlation-induced information overcounting.

Table 9. Sensitivity to cross-correlation coefficient on the CT model.

Correlation Coefficient	EIF-Inexact RMSE	Measurement Fusion RMSE	KalmanNet RMSE	ACIFNet RMSE
0.00	8.15	8.02	7.14	5.88
0.25	9.36	9.22	7.48	5.97
0.50	11.52	11.38	8.05	6.18
0.75	14.21	14.58	9.73	7.15
1.00	18.65	19.30	12.86	9.84

Table 10. Sensitivity to common disturbance strength on the CT model.

Common Disturbance Coefficient	EIF-Inexact RMSE	Measurement Fusion RMSE	KalmanNet RMSE	ACIFNet RMSE
1	10.74	10.51	7.52	5.92
2	11.52	11.38	8.05	6.18
4	13.86	14.12	9.48	7.02
8	17.34	17.89	11.36	8.25

Table 11 further examines the influence of the number of sensors. Adding more sensors generally improves localization accuracy because more observations become available. However, the improvement becomes smaller when the added sensors are affected by similar common disturbances. The inference time increases with the number of sensors because the Transformer processes more sensor tokens, but the cost remains moderate for the tested sparse sensor network.

Table 11. Sensitivity to the number of sensors on the CT model.

Number of Sensors	EIF-Inexact RMSE	Measurement Fusion RMSE	KalmanNet RMSE	ACIFNet RMSE	ACIFNet Time/ms
2	11.52	11.38	8.05	6.18	5.60
3	10.84	10.76	7.78	5.74	6.92
4	10.31	10.25	7.55	5.51	8.36
6	10.05	9.98	7.33	5.33	12.70

These sensitivity results indicate that ACIFNet is most effective when reliable but cross-correlated sensor information still exists. When the common disturbance is too strong or when most sensors become severely degraded, the amount of useful independent information is limited, and the performance gain naturally decreases. This observation is consistent with the robustness results in Section 6.6 and clarifies the applicable range of the proposed method.

7. Conclusions

This paper proposed ACIFNet, an adaptive correlation information filter network for multi-sensor fusion localization of intelligent vessels under unknown cross-correlated

measurement noises. By preserving the recursive structure of the extended information filter, ACIFNet converts local sensor estimates into information-domain contributions and uses a Transformer-based network to learn adaptive fusion weights, thereby compensating for unknown inter-sensor correlations without explicitly estimating the full cross-correlation covariance matrix.

Experiments on the CV, CT, and 3-DOF vessel motion models, as well as a real-world restricted-waterway dataset, demonstrate that ACIFNet achieves lower and more stable localization errors than EIF-inexact fusion, measurement fusion, and KalmanNet. The robustness, ablation, runtime, and sensitivity results further confirm that adaptive information weighting is the main source of improvement, while temporal information features and the Transformer module enhance fusion stability. ACIFNet remains effective under different cross-correlation strengths, sensor reliability mismatches, and sensor numbers, although its advantage decreases when common disturbances become severe or reliable sensor information is insufficient.

Several limitations remain. The current experiments mainly focus on two-sensor fusion, and larger heterogeneous sensor networks and asynchronous multi-rate scenarios require further validation. The cross-correlated noise is partly simulated using shared disturbances, so more real-world datasets with natural sensor correlations are needed. In addition, the learned weights are constrained for numerical consistency, but this does not provide a formal global stability guarantee. ACIFNet also relies on supervised training and introduces additional computational cost compared with conventional EIF.

Future work will extend ACIFNet to larger-scale sensor networks, asynchronous fusion, missing-measurement scenarios, online uncertainty calibration, and lightweight embedded deployment in complex marine and port environments.

Author Contributions: Conceptualization, L.Y., Z.X. and B.M.; methodology, L.Y. and W.Z.; software, W.Z.; validation, L.Y. and J.G.; formal analysis, L.Y.; investigation, B.M.; resources, B.M. and J.G.; data curation, Z.X. and L.Y.; writing—original draft preparation, W.Z. and L.Y.; writing—review and editing, L.Y. and Z.X.; visualization, J.G.; supervision, D.K.; project administration, L.Y. and B.M.; funding acquisition, L.Y. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by the National Natural Science Foundation of China under Grant 62301136, the S&T Program of Hebei under Grant 242G1602Z, and the Central Guidance for Local Science and Technology Development Funds Project under Grant 254Z1603G.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Conflicts of Interest: Authors Lei Yan and Bo Meng were employed by the company Hebei Port Group Co., Ltd. Authors Zhixin Xia and Junli Ge were employed by the company Shulian Technologies Limited, Hebei Port Group Co., Ltd. Author Deming Kong was employed by the company Hebei Yansoft Information System Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Pastore, T.; Djapic, V. Improving autonomy and control of autonomous surface vehicles in port protection and mine countermeasure scenarios. *J. Field Robot.* **2010**, *27*, 903–914. [CrossRef]
2. Naeem, W.; Irwin, G.W.; Yang, A. COLREGs-based collision avoidance strategies for unmanned surface vehicles. *Mechatron* **2012**, *22*, 669–678. [CrossRef]

3. Ma, H.; Smart, E.; Ahmed, A.; Brown, D. Radar image-based positioning for USV under GPS denial environment. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 72–80. [CrossRef]
4. Abu Bakr, M.; Lee, S. Distributed multisensor data fusion under unknown correlation and data inconsistency. *Sensors* **2017**, *17*, 2472. [CrossRef] [PubMed]
5. Julier, S.J.; Uhlmann, J.K. A non-divergent estimation algorithm in the presence of unknown correlations. In Proceedings of the American Control Conference, Albuquerque, NM, USA, 6 June 1997; pp. 2369–2373. [CrossRef]
6. Clark, D.; Julier, S.; Mahler, R.; Ristic, B. Robust multi-object sensor fusion with unknown correlations. In Proceedings of the Sensor Signal Processing for Defence, London, UK, 29–30 September 2010; pp. 1–5. [CrossRef]
7. Bar-Shalom, Y.; Campo, L. The effect of the common process noise on the two-sensor fused-track covariance. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *AES-22*, 803–805. [CrossRef]
8. Simon, D. *Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
9. Revach, G.; Shlezinger, N.; Ni, X.; Escoriza, A.L.; van Sloun, R.J.G.; Eldar, Y.C. KalmanNet: Neural network aided Kalman filtering for partially known dynamics. *IEEE Trans. Signal Process.* **2022**, *70*, 1532–1547. [CrossRef]
10. Zhang, H.; Wang, W.; Liu, C.; Zhang, X. Transformer aided adaptive extended Kalman filter for vehicle mass estimation. *Processes* **2023**, *11*, 887. [CrossRef]
11. Shen, S.; Wang, Z.; Zhang, Y. KalmanFormer: Using Transformer to model the Kalman gain in Kalman filters. *Front. Neurorobot.* **2025**, *18*, 1460255. [CrossRef] [PubMed]
12. Li, J.; Sun, R.; Wang, Y.; Ochieng, W.Y. A robust time synchronization algorithm for GNSS/IMU integrated navigation in urban environments. *Meas. Sci. Technol.* **2025**, *36*, 036302. [CrossRef]
13. Li, D.; Zhang, F.; Rong, W.; Yue, C.; Zhang, Y.; Liang, Y.; Ren, J. Robust localization algorithm for micromanipulation targets under complex interference conditions. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 23959–23969. [CrossRef]
14. Liang, Z.; Liu, L.; Yang, J.; Yu, J.; Zhang, S.; Zhang, X.; Wang, Y. AUV cooperative localization method based on M-H resampling and state covariance matrix optimized particle filter. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 8513314. [CrossRef]
15. Wang, J.; Li, R.; Lu, X.; Jiang, J.; Wan, P.; Chu, H. Risk-averse robustness-based intelligent environment-adaptive GNSS/INS tightly-coupled positioning method. *IEEE Trans. Veh. Technol.* **2026**, *Early Access*. [CrossRef]
16. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]
17. Julier, S.J.; Uhlmann, J.K. New extension of the Kalman filter to nonlinear systems. In *Signal Processing, Sensor Fusion, and Target Recognition VI*; SPIE: Bellingham, WA, USA, 1997; Volume 3068, pp. 182–193. [CrossRef]
18. Mutambara, A.G.O. *Decentralized Estimation and Control for Multisensor Systems*; CRC Press: Boca Raton, FL, USA, 1998.
19. Tang, Q.; Duan, Z.; Li, X.R. Modeling and analysis of correlation of local estimation errors for distributed BLUE fusion. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 6995–7011. [CrossRef]
20. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017; pp. 5999–6009.
21. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
22. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
23. Li, X.R.; Jilkov, V.P. Survey of maneuvering target tracking. Part I: Dynamic models. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 1333–1364. [CrossRef]
24. Fossen, T.I. *Handbook of Marine Craft Hydrodynamics and Motion Control*; John Wiley & Sons: Chichester, UK, 2011.
25. Chung, D.; Kim, J.; Lee, C.; Kim, J. Pohang canal dataset: A multimodal maritime dataset for autonomous navigation in restricted waters. *Int. J. Robot. Res.* **2023**, *42*, 1104–1114. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.