

AI-Enhanced Sensor Fusion: Bridging Traditional and Neural Network Filters in Real-Time Localization

Aaryaa Padhyegurjar
RPTU Kaiserslautern-Landau
Kaiserslautern, Germany
aaryaa_shekharpadhyegurjar@dfki.de

Dennis Salzmann
DFKI GmbH
Kaiserslautern, Germany
dennis.salzmann@dfki.de

Hans D. Schotten
RPTU Kaiserslautern-Landau
Kaiserslautern, Germany
schotten@rptu.de

Abstract—In the context of digital twins and industrial IoT systems, robust sensor fusion is essential for accurate real-time localization. This paper presents a comparative evaluation of three sensor fusion strategies for indoor 2D localization: a classical Extended Kalman Filter (EKF), a hybrid architecture combining the filter with a Recurrent Neural Network (RNN) for error compensation, and a pure AI model that learns the fusion process directly from multi-sensor sequences.

Experimental results show that the hybrid approach consistently achieves the best trade-off between accuracy and real-time responsiveness, outperforming both the classical and the purely AI-based alternative.

I. INTRODUCTION

Accurate real-time localization is a key requirement for autonomous robots, automated guided vehicles (AGVs), and Industry 4.0 systems. With the rise of digital twins and pervasive IoT infrastructures, reliable localization is essential to ensure safety and efficiency in dynamic indoor environments.

Sensor fusion addresses this challenge by combining heterogeneous inputs such as inertial measurement units, ultrasonic beacons, ultra-wideband (UWB) tags, and optical trackers [1], [2]. The Extended Kalman Filter (EKF) has long been the standard approach for sensor fusion due to its recursive efficiency, probabilistic structure, and interpretability [3], [4]. However, EKF performance is limited by its reliance on accurate models, Gaussian noise assumptions, and synchronous measurements.

Recent advances in artificial intelligence offer alternatives. Recurrent Neural Networks (RNNs), particularly Gated Recurrent Units (GRUs), can learn temporal patterns and nonlinear sensor dynamics directly from data [5], [6]. By bypassing explicit motion models, such methods promise adaptability to complex environments. However, they typically demand extensive datasets and exhibit potential stability issues.

Hybrid methods seek to combine the strengths of both paradigms. By augmenting a Kalman filter with a learned correction or adaptive mechanism, they aim to retain stability and interpretability while improving accuracy in noisy, asynchronous conditions [7]–[10]. Early studies indicate their promise, but few works provide systematic real-world comparisons of classical, hybrid, and fully AI-based pipelines.

The main contribution of this work is an experimental comparison of classical, hybrid, and AI-based sensor fusion pipelines implemented on the same dataset and platform. The paper is organized as follows: Section II presents the framework architecture, Section III details the methodology, Section IV describes the experimental setup, Section V reports the results, Section VI discusses the findings, and Section VII concludes the paper.

II. ARCHITECTURE

A. Hardware Setup

The experimental platform was deployed in the 6G Future Factory Lab at DFKI Kaiserslautern as part of the Twin4Trucks project. Multiple positioning systems were mounted on an automated guided vehicle (AGV), providing heterogeneous and asynchronous sensor data.

The ground truth was provided by a motion capture system consisting of 24 infrared cameras, delivering sub-millimeter accuracy. Three commercial indoor localization technologies were mounted on the AGV for the experiments:

- 1) Ultra-Wideband (UWB): RF-based tags and anchors measuring time-of-flight to provide 3D position estimates.
- 2) Ultrasonic: Fixed beacons and a mobile tag providing trilateration-based positions, with optional IMU-based motion estimates.
- 3) Infrared: Sweep-based system delivering six-degree-of-freedom position and orientation tracking.

Such heterogeneous sensor setups are increasingly common in industrial and robotic localization tasks, where combining ultrasonic and infrared systems has shown complementary benefits [11].

B. Software Setup

To handle multiple sensor streams and maintain modularity, the system was implemented as a containerized software pipeline. Two core services were developed:

- 1) **SensorFusionNode**: Implements the three fusion pipelines:

- a) Classical Extended Kalman Filter (EKF) for baseline comparison.
 - b) Hybrid fusion with error compensation, which learns residual corrections.
 - c) Standalone AI-based fusion, which directly predicts fused positions.
- 2) **MockSensorNode**: Replays previously recorded sensor logs with original timestamps to emulate real-time asynchronous behavior. Gaussian noise was added to preserve variability and test robustness.

All components communicate through a standardized message format defined in Protocol Buffers (Protobuf), ensuring consistent data representation across the fusion pipelines. This architecture enables reproducible evaluation of classical, hybrid, and AI-based approaches under identical runtime conditions.

III. METHODOLOGY

This work evaluates three complementary sensor fusion strategies: a classical Extended Kalman Filter, a hybrid fusion approach, and an AI-based fusion model. All methods are implemented within the same software framework to ensure comparability under real-time conditions.

A. Classical Extended Kalman Filter

The Extended Kalman Filter (EKF) serves as the baseline due to its widespread use in robotics and real-time localization. The state vector is defined as

$$\mathbf{x} = [x \quad y \quad \theta]^T, \quad (1)$$

where (x, y) denote position and θ is the orientation (yaw). Control inputs are the linear and angular velocities $\mathbf{u} = [v, \omega]^T$, which in this work are derived from ground truth trajectories to maintain consistency, as opposed to onboard odometry, from which these values are typically obtained.

The EKF prediction step propagates the state using the nonlinear motion model $f(\mathbf{x}, \mathbf{u})$, while the update step incorporates asynchronous sensor measurements through the observation model $h(\mathbf{x})$. A conservative covariance inflation strategy is applied across all modalities, ensuring filter stability under heterogeneous and noisy conditions [12].

B. Hybrid EKF with GRU-Based Error Compensation

To address limitations of purely model-based filtering, the EKF is augmented with a Gated Recurrent Unit (GRU) network. After each EKF update, the fused position estimate $\hat{\mathbf{x}}_k$ (from Eq. 1) is stored in a sliding window of 50 time steps. This sequence forms the GRU input, which predicts the residual error $\Delta \mathbf{e}_k$ between the EKF output and ground truth. The compensated position is obtained as given in Eq. 2:

$$\hat{\mathbf{x}}_k^{comp} = \hat{\mathbf{x}}_k - \Delta \mathbf{e}_k \quad (2)$$

GRU architectures were optimized using Optuna hyperparameter search, with evaluation metrics including mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R^2). All training and validation runs were tracked in MLflow.

C. AI-Based Fusion

The third approach removes the filter framework entirely. Instead, a GRU model is trained to directly predict the fused position from sequences of asynchronous sensor readings. Each input consists of a sliding window of 10 multi-sensor measurements, standardized and aligned in time. The GRU outputs the estimated current position (x, y) without requiring motion models, control vectors, or covariance matrices.

This method provides a lightweight inference pipeline but requires larger training datasets to generalize. While less stable in edge cases, it serves as a benchmark for assessing the potential of fully end-to-end AI-based sensor fusion under real-time constraints.

TABLE I. COMPARISON OF GRU-BASED ARCHITECTURES FOR HYBRID AND AI-BASED FUSION MODELS

No.	Layer Type	Hybrid Fusion	AI-Based Fusion
1	GRU	256	256
2	Dropout	0.3	0.3
3	GRU	128	128
4	Dropout	0.1	-
5	GRU	64	-

Table I summarizes the GRU-based architectures used in the Hybrid and AI-Based fusion models. Both architectures consist of stacked GRU and dropout layers with varying depths, as shown. Each model is followed by an output layer of type Dense. All GRU layers use the ReLU activation function.

IV. EXPERIMENTAL SETUP

A. Data Collection

Experiments were conducted using an Automated Guided Vehicle (AGV) equipped with three different positioning systems, as described in Section II. The AGV was driven along trajectories designed to cover a variety of motion patterns, including straight lines, curves, and acceleration phases, in order to expose the fusion algorithms to diverse kinematic conditions.

B. Evaluation Metrics

Performance was assessed using complementary error metrics that capture different aspects of localization quality. Mean Absolute Error (MAE) reflects the average deviation from ground truth, while Root Mean Squared Error (RMSE) emphasizes the impact of larger deviations. The coefficient of determination (R^2) was used to quantify how well predicted trajectories explain the variance of the ground truth.

However, as the primary measure, we focused on the *Mean Euclidean Error*, which directly reflects trajectory-level accuracy in two dimensions and is also used in the visual comparison of methods. This metric was used for the final comparisons.

In addition to accuracy, inference time per update was measured to evaluate real-time feasibility. A threshold of 100 ms was applied, representing the maximum acceptable delay for continuous AGV localization.

V. RESULTS

A. Trajectory Comparison

Fig. 1 shows a representative trajectory segment for all three fusion methods compared against the ground truth.

- The **EKF** baseline exhibits noticeable drift, particularly on curved paths.
- The **Hybrid approach** remains closest to ground truth, though occasional overcompensation is observed.
- The **AI-based approach** tracks overall motion well but produces jittery behavior in cornering sections.

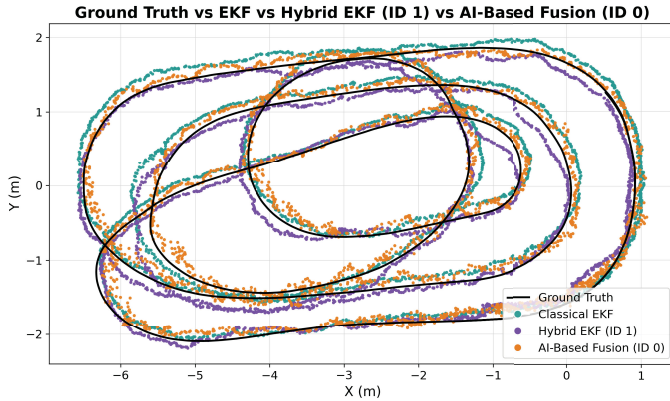


Fig. 1. Trajectory comparison of classical, hybrid and AI-based approaches against ground truth.

B. Error Heatmaps and CDF Analysis

Fig. 2 illustrates spatial error distributions across the test environment. The classical EKF shows concentrated bias in curved regions, while the Hybrid filter distributes errors more evenly and reduces overall magnitude. The AI-based model demonstrates competitive accuracy but introduces localized instabilities.

Cumulative Distribution Functions (CDFs) of Euclidean error are shown in Fig. 3. The Hybrid filter achieves the most favorable error distribution, with a clear left shift indicating that lower errors occur more frequently. The AI-based model approaches Hybrid performance in mid-range errors but is less reliable at extreme values. The EKF remains the most stable but suffers from consistently higher error levels.

VI. DISCUSSION

The experimental results highlight the complementary strengths and weaknesses of classical, hybrid, and AI-based fusion approaches.

The **classical EKF** demonstrated the expected robustness and computational efficiency. Its recursive structure guarantees stability under diverse runtime conditions. However, the EKF struggled with curved motion segments and sensor asynchrony, resulting in higher overall error compared to learning-based approaches.

The **hybrid EKF** consistently outperformed both alternatives in terms of accuracy and robustness. By leveraging

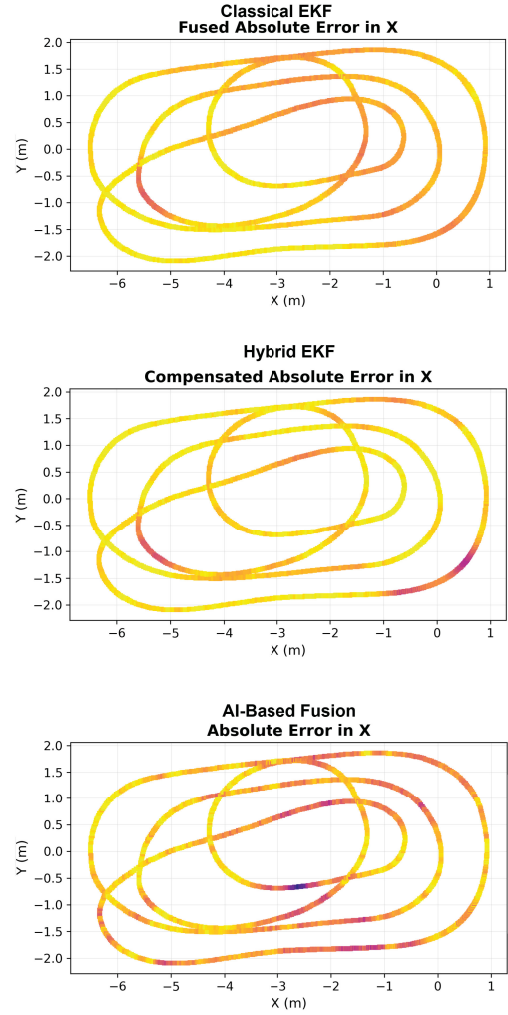


Fig. 2. Error heatmaps for all three approaches. Darker color indicates higher error.

the EKF's probabilistic structure and the GRU's capacity to learn residual dynamics, the hybrid method achieved the most favorable error distribution while remaining real-time capable. Occasional overcompensation was observed, indicating that learned corrections may amplify errors when the filter state is already close to the true trajectory.

The **AI-based GRU** fusion approach showed competitive accuracy and in some cases rivaled the hybrid method. Its advantage lies in bypassing the need for motion models or noise parameter tuning, making it attractive for deployment in environments where system dynamics are poorly understood. However, it also exhibited higher variance, particularly in edge cases, suggesting limited generalization compared to the hybrid method.

Overall, the results confirm that combining model-based and learning-based components provides a practical balance between interpretability, stability, and adaptability. This trade-off is summarized in Fig. 4, which compares Mean Euclidean

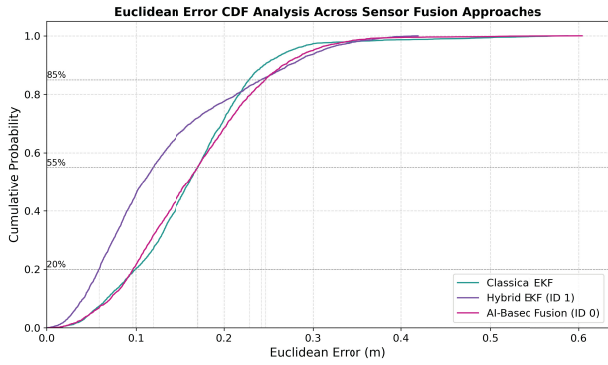


Fig. 3. CDF of Euclidean error across all methods.

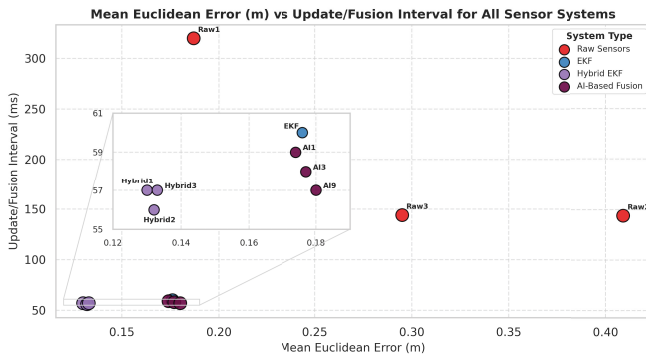


Fig. 4. Comparison of Mean Euclidean Error versus fusion/update interval across the classical, hybrid and AI-based approaches. Lower-left indicates the most desirable trade-off between accuracy and latency.

Error against fusion/update interval. The plot illustrates that the hybrid EKF is consistently closest to the ideal low-error, low-latency region, making it the most suitable approach for safety-critical real-time applications such as AGV localization in industrial environments.

While a detailed computational complexity analysis was beyond the scope of this work, all methods maintained inference times below 100 ms per update, fulfilling real-time localization requirements.

VII. CONCLUSION AND FUTURE WORK

This paper presented a comparative study of classical, hybrid, and AI-based sensor fusion pipelines under identical experimental conditions. The results showed that the hybrid fusion approach consistently achieved the best trade-off between accuracy and responsiveness, confirming that learned residuals can effectively complement model-based estimation. The classical EKF remained the most computationally efficient

but suffered from higher error. On the other hand, the AI-based approach produced competitive results but lacked robustness across all motion scenarios.

Future work can explore deeper integration of learning into model-based filtering, such as adaptive tuning of process and measurement covariances (Q and R) or reliability-aware weighting of individual sensor streams. In addition, multi-modal neural architectures and attention mechanisms could further improve adaptability.

ACKNOWLEDGMENT

The authors acknowledge the financial support by the German *Federal Ministry of Research, Technology and Space (BMFTR)* within the project Twin4Trucks {13IK010F}.

REFERENCES

- [1] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1566253511000558>
- [2] S. Wang and N. Ahmad, "A comprehensive review on sensor fusion techniques for localization of a dynamic target in gps-denied environments," *IEEE Access*, vol. PP, pp. 1–1, 01 2024.
- [3] G. Welch and G. Bishop, "An introduction to the kalman filter," *Proc. Siggraph Course*, vol. 8, 01 2006.
- [4] M. Grewal and A. Andrews, "Kalman filtering: theory and practice using matlab," *New York: John Wiley and Sons*, vol. 14, 01 2001.
- [5] C. Wang, W. Du, Z. Zhu, and Z. Yue, "The real-time big data processing method based on lstm or gru for the smart job shop production process," *Journal of Algorithms & Computational Technology*, vol. 14, p. 1748302620962390, 2020. [Online]. Available: <https://doi.org/10.1177/1748302620962390>
- [6] M. Hussain, M. O'Nils, J. Lundgren, and S. J. Mousavirad, "A comprehensive review on deep learning-based data fusion," *IEEE Access*, vol. 12, pp. 180 093–180 124, 2024.
- [7] S. Kim, I. Petrunin, and H.-S. Shin, "A review of kalman filter with artificial intelligence techniques," in *2022 Integrated Communication, Navigation and Surveillance Conference (ICNS)*, 2022, pp. 1–12.
- [8] G. Revach, N. Shlezinger, R. J. G. van Sloun, and Y. C. Eldar, "Kalmannet: Data-driven kalman filtering," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 3905–3909.
- [9] M. Jahja, D. Farrow, R. Rosenfeld, and R. J. Tibshirani, "Kalman filter, sensor fusion, and constrained regression: Equivalences and insights," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/b522259710151f8cc7870b970b4e0930-Paper.pdf
- [10] M. A. Lee, B. Yi, R. Martín-Martín, S. Savarese, and J. Bohg, "Multimodal sensor fusion with differentiable filters," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 444–10 451.
- [11] E. Aparicio-Esteve, J. Ureña, A. Hernández, and J. M. Villadangos, "Combined infrared-ultrasonic positioning system to improve the data availability," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 25 152–25 164, 2023.
- [12] S. Reece and S. Roberts, "Generalised covariance union: A unified approach to hypothesis merging in tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 207–221, 2010.