This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Physics-Informed Residual Learning Method for Real-Time 5-DoF Magnetic Localization in Capsule Endoscopy

Miaozhang Shen, *Graduate Student Member, IEEE*, Shuxiang Guo, *Fellow, IEEE*, Zixu Wang, *Member, IEEE*, Yuyue Yang, Chunying Li, *Senior Member, IEEE*, Mingchao Ding, and Bing Wang

***Abstract*—Real-time localization is a requirement for the clinical application of magnetically actuated wireless capsule endoscopy. This study proposes a physics-informed deep learning framework to address the model-plant mismatch inherent in analytical methods and the data requirements of end-to-end learning. A dual-branch residual network, physics-informed residual network, computes corrections for the magnetic dipole model. Two operational paradigms are evaluated: a lookup-table-based approach and an end-to-end optimization approach. Experiments indicate that the framework achieves a mean position error of 0.69 $\pm$ 0.29 mm and a mean orientation error of 0.70 $\pm$ 0.37°, utilizing less than 20% of the calibration configurations required by data-driven baselines. Validation includes an ex vivo signal propagation study through biological tissue and compatibility tests with an anti-interference filter, supplemented by a preliminary in vivo animal experiment. The results provide a methodology for sim-to-real transfer in robotic capsule tracking systems.**

***Index Terms*—Anti-interference integration, ex vivo validation, magnetic capsule endoscopy, physics-informed neural network, sim-to-real.**

Received 8 April 2026; accepted 22 April 2026. This work was supported in part by the Shenzhen “Medical Engineering Integration” Special Support Program Project under Grant F-2024-Z99-502365 and in part by the High level of special funds from Southern University of Science and Technology under Grant G03034K003. Paper no. TII-26-3112. *(Corresponding authors: Shuxiang Guo; Zixu Wang.)*

This work involved animal subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Animal Ethics Committee of Southern University of Science and Technology under Application No. SUSTech-SL2025042703 and IACUC Issue No. SUSTech-JY202507033, and performed in line with the Guiding Principles in the Care and Use of Animals.

Miaozhang Shen is with the Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518000, China, and also with the Advanced Institute for Ocean Research, Southern University of Science and Technology, Shenzhen 518000, China (e-mail: shenmz2024@mail.sustech.edu.cn).

Zixu Wang is with the Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518000, China (e-mail: wangzx3@sustech.edu.cn).

Shuxiang Guo is with the Department of Electronic and Electrical Engineering, and the Advanced Institute for Ocean Research, Southern University of Science and Technology, Shenzhen 518000, China, and also with the Aerospace Center Hospital, School of Life Science and the Key Laboratory of Convergence Medical Engineering System and Healthcare Technology, Ministry of Industry and Information Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: guo.shuxiang@sustech.edu.cn).

Yuyue Yang is with the School of Electronic Science and Engineering, South China Normal University, Guangdong Province Key Lab of Chip and Integration Technology, Foshan 528000, China.

Chunying Li is with the Advanced Institute for Ocean Research, Southern University of Science and Technology, Shenzhen 518000, China.

Mingchao Ding and Bing Wang are with the Department of Peripheral Vascular Intervention, Aerospace Center Hospital, School of Life Science, Beijing Institute of Technology, Beijing 100081, China.

Digital Object Identifier 10.1109/TII.2026.3688686

I. Introduction

WIRELESS capsule endoscopy characterizes an established noninvasive modality, which provides an accessible alternative for gastrointestinal diagnostics [1], [2], [3]. Advancements in micro-robotics are transitioning these devices from passive imaging tools toward active robotic platforms capable of executing in vivo tasks, which include targeted drug delivery, tissue biopsy, and minimally invasive surgical procedures [4], [5], [6], [7]. The realization of such therapeutic potential depends on the capability to determine the 5-degree-of-freedom pose, which encompasses the translation and orientation of the capsule within the human anatomy [8], [9], [10]. Permanent magnet tracking characterizes a preferred solution for these applications because it functions without internal power supplies and maintains signal transmission through biological media without line-of-sight requirements . However, inaccurate localization renders active control unfeasible and increases the risk of tissue trauma, which constrains the clinical translation of these robotic systems [11], [12].

Current localization strategies predominantly follow two distinct paradigms, which involve model-based optimization and data-driven learning. The first paradigm relies on fitting analytical physical models, which include the magnetic dipole model to sensor measurements via iterative algorithms, such as Levenberg–Marquardt (LM) [13], [14]. The performance of this paradigm is limited by model-plant mismatch, which arises from idealized dipole assumptions and ambient noise. Research by Fu and Guo [15] indicates that such discrepancies lead to localization errors in the near-field regions, which are requisite for surgical operations. Furthermore, the accuracy of analytical models is compromised by inevitable manufacturing tolerances that include deviations in magnet dimensions and magnetization

TABLE I
STRUCTURAL COMPARISON WITH RELATED SIM-TO-REAL MAGNETIC LOCALIZATION PARADIGMS

Feature	ResNet-LM [22]	AMagPoseNet [18]	MobilePoseNet [11]	PIRNet (Ours)
Correction level	Pose	Pose	Sensor/Pose	Field (ΔB)
Sim-to-real strategy	Data augment.	Multistage	Sim pretrain	Sim pretrain + fine-tune
Physics regularization	None	None	None	$\mathcal{L}_{\text{p-div}}$
Forward model reuse	No	No	No	Yes (differentiable)
Dual deployment mode	No	No	No	Yes (LUT / in-the-Loop)
Disentangled error sources	No	No	Partial	Yes (dual-branch)

intensity which are inherent in miniaturized components [6]. While complex numerical integration models improve accuracy, they introduce high computational costs, which hinder real-time performance during dynamic tracking [16], [17].

The second paradigm utilizes deep learning architectures to map raw sensor readings to the capsule pose [18], [19], [20]. Although these methods capture nonlinear characteristics, they are data-intensive and require extensive labeled real-world datasets acquired through complex hardware configurations [11], [21], [22]. Sebkhi et al. [23] quantified this calibration burden by demonstrating that the generation of 1.7 million samples for a 5-D positional stage requires six months of operation. Moreover, pure neural network models operate as black boxes, which often violate Maxwell equations and produce physically inconsistent field gradients [16], [24]. This lack of physical grounding results in gradient prediction errors, which degrade the performance of force-based control in magnetic navigation systems. In addition, such models demonstrate limited robustness when encountering out-of-distribution conditions, which include dynamic distortions from ferromagnetic surgical instruments [25].

Existing sim-to-real magnetic localization methods can be categorized along two design axes: *what is learned* and *how physics is enforced*. Pose-level correction methods, such as the ResNet-LM fusion by Guo et al. [22] and the multistage training pipeline of AMagPoseNet [18], learn a direct mapping from sensor readings to pose corrections, treating the physical model as a black-box initialization or data augmentation source. Calibration subnetwork approaches, such as the denoising modules in MobilePoseNet [11], learn sensor-level noise patterns but do not model the systematic discrepancy between the analytical forward model and the physical reality. In contrast, the framework proposed herein operates at the *field level:* physics-informed residual network (PIRNet) learns a spatially structured residual field $\Delta\mathbf{B}$ that explicitly corrects the forward model itself, rather than its output pose. This field-level formulation enables two capabilities absent in prior work: 1) the corrected forward model can be reused as a differentiable physics layer within arbitrary downstream solvers (the PIRNet-in-the-Loop paradigm), and 2) the learned residual field can be regularized by Maxwell-derived spatial constraints ($\mathcal{L}_{\text{p-div}}$), which is not applicable to pose-level corrections that lack spatial structure. Table I summarizes these structural distinctions.

Real-world surgical environments introduce the requirement for system-level robustness against dynamic ferromagnetic interference. Surgical instruments composed of soft-magnetic materials induce nonlinear distortions in the local magnetic field, which confound traditional localization solvers [16], [25]. While hybrid approaches attempt to bridge these paradigms via sim-to-real transfer, existing research lacks a systematic framework that explicitly disentangles local sensor-level distortions from global pose-dependent deviations [18], [22]. There exists a necessity for localization methods, which maintain physical consistency and modular compatibility with downstream filtering architectures while minimizing the real-world data dependence.

This article proposes the PIRNet to address the model-plant mismatch and calibration burden in magnetic tracking. The architecture disentangles error sources through a dual-branch structure, which computes a physically plausible correction to the idealized dipole model. The main contributions of this article are summarized as follows.

1) A PIRNet architecture is proposed, which utilizes a dual-branch encoder to decouple local sensor distortions and systematic pose-dependent deviations while enforcing an in-plane pseudodivergence regularization.
2) Two operational paradigms are formulated and compared: a lookup table (LUT)-based approach for high-speed edge computing and an end-to-end optimization approach for enhanced accuracy.
3) A data-efficient training strategy is demonstrated, achieving submillimeter accuracy with less than 20% of the configurations required by conventional data-driven methods, thereby reducing the spatial calibration burden.
4) The robustness of the framework is assessed through a progressive experimental protocol, including ex vivo tissue penetration studies and integration tests with anti-interference architectures under representative preclinical conditions.

II. METHODOLOGY

This section presents the proposed physics-informed deep learning framework for magnetic localization. First, the foundational magnetic dipole forward model is briefly introduced. Next, the core architecture of PIRNet is presented. Finally, the two operational paradigms derived from PIRNet, namely, PIRNet-LUT and PIRNet-in-the-Loop, are described in detail.

A. Magnetic Dipole Forward Model

The magnetic field generated by a permanent magnet can be approximated by the magnetic dipole model, which has been established as a foundational method for wireless capsule endoscopy (WCE) localization [26]. The magnetic flux density vector $\mathbf{B}$ at an observer point $\mathbf{r}$ generated by a magnet centered at $\mathbf{s}$ with a magnetic moment $\mathbf{m}$ can be expressed as

$$\mathbf{B}(\mathbf{r},\mathbf{s},\mathbf{m}) = \frac{\mu_0}{4\pi}\left(\frac{3(\mathbf{m}\cdot\mathbf{R})\mathbf{R}}{||\mathbf{R}||^5} - \frac{\mathbf{m}}{||\mathbf{R}||^3}\right) \quad (1)$$

where $\mathbf{R} = \mathbf{r} - \mathbf{s}$ is the relative position vector, and μ_0 is the magnetic permeability of free space. Given the pose of the magnet (position $\mathbf{s}$ and orientation, which determines $\mathbf{m}$) and the positions of an array of N sensors $\{\mathbf{r}_i\}_{i=1}^N$, this forward model can predict the theoretical magnetic field matrix $\mathbf{B}_{\text{sim}} \in$

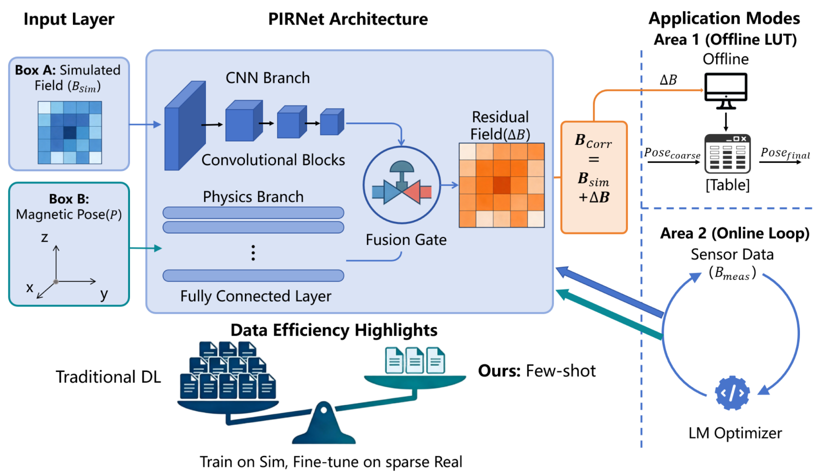


Fig. 1. Architecture of the proposed PIRNet.

$\mathbb{R}^{N\times 3}$. However, as discussed, this model suffers from inherent inaccuracies. The objective is to learn a correction to this model.

B. PIRNet Architecture

The core of the framework is the PIRNet, a neural network specifically designed to learn the residual correction $\Delta\mathbf{B}$ between the real-world magnetic field and the predictions of the dipole model. The final high-fidelity field prediction is then $\mathbf{B}_{\text{corr}} = \mathbf{B}_{\text{sim}} + \Delta\mathbf{B}$. The architecture, illustrated in Fig. 1, is detailed below. The PIRNet features a dual-branch encoder to process multimodal inputs, a gated fusion module to intelligently combine features, and an output head to generate the final correction.

1) Dual-Branch Feature Encoder:

a) *Residual CNN branch:* This branch processes the simulated magnetic field matrix $\mathbf{B}_{\text{sim}} \in \mathbb{R}^{5\times 5\times 3}$ to capture local sensor-level distortion patterns. It consists of a series of standard residual blocks [27], where shortcut connections with 1×1 convolutions are used to handle dimensional changes. This branch outputs a local feature map $\mathbf{F}_{\text{cnn}} \in \mathbb{R}^{5\times 5\times 64}$.

b) *Physics encoder branch:* This MLP-based branch processes the 6-D magnet pose vector $\mathbf{p} \in \mathbb{R}^{6}$ to model systematic pose-dependent model deviations. The input vector is passed through several fully connected layers to generate a 64-D feature vector, which is then spatially broadcast to form a physics feature map $\mathbf{F}_{\text{phy}} \in \mathbb{R}^{5\times 5\times 64}$.

2) Context-Aware Gated Fusion: The two feature maps are combined using a context-aware gating mechanism. A contextual feature map, $\mathbf{F}_{\text{context}}$, is first generated by combining $\mathbf{F}_{\text{cnn}}$ and $\mathbf{F}_{\text{phy}}$. A dynamic gating weight map $\sigma \in [0,1]^{5\times 5\times 64}$ is then computed from $\mathbf{F}_{\text{context}}$ using a convolutional layer and a Sigmoid function. The final fused feature $\mathbf{F}_{\text{fusion}}$ is calculated as

$$\mathbf{F}_{\text{fusion}} = \sigma \otimes \mathbf{F}_{\text{cnn}} + (1-\sigma) \otimes \mathbf{F}_{\text{phy}} \tag{2}$$

where $\otimes$ denotes elementwise multiplication.

3) Output Head: The fused feature map $\mathbf{F}_{\text{fusion}}$ is then passed through a final convolutional layer (1×1 Conv, three channels), which serves as the output head. This layer decodes the high-level features into the final three-channel residual correction matrix, $\Delta\mathbf{B} \in \mathbb{R}^{5\times 5\times 3}$.

4) Physics-Informed Training Objective: To train the PIRNet, a composite loss function is designed to simultaneously encourage fitting to ground-truth data while ensuring spatial smoothness and physical plausibility of the predictions. The total loss $\mathcal{L}$ for a given sample is defined as a weighted sum of a data-fitting term and a physics-inspired regularization term

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda \cdot \mathcal{L}_{\text{p-div}} \tag{3}$$

where λ is a hyperparameter that balances the two terms.

The data-fitting loss, $\mathcal{L}_{\text{data}}$, measures the squared L_2 norm of the difference between the final corrected field $\mathbf{B}_{\text{corr}}$ and the real-world ground-truth field $\mathbf{B}_{\text{real}}$

$$\mathcal{L}_{\text{data}} = ||\mathbf{B}_{\text{corr}} - \mathbf{B}_{\text{real}}||_2^2 = ||(\mathbf{B}_{\text{sim}} + \Delta\mathbf{B}) - \mathbf{B}_{\text{real}}||_2^2. \tag{4}$$

Ideally, the predicted magnetic field should strictly adhere to Maxwell's equations, specifically the divergence-free constraint ($\nabla\cdot\mathbf{B} = 0$). Since the analytical dipole model field $\mathbf{B}_{\text{sim}}$ is inherently divergence-free, this constraint should theoretically apply directly to the learned residual field $\Delta\mathbf{B}$, that is, $\frac{\partial\Delta B_x}{\partial x} + \frac{\partial\Delta B_y}{\partial y} + \frac{\partial\Delta B_z}{\partial z} = 0$.

However, the physical limitation of the hardware setup is explicitly acknowledged: acquiring the full 3-D spatial gradient (specifically, the z-axis partial derivative $\frac{\partial\Delta B_z}{\partial z}$) is mathematically ill-posed when utilizing a single 2-D planar sensor array ($Z = 0$ plane). To address this limitation without discarding the valuable physical prior, a relaxed in-plane pseudodivergence penalty, $\mathcal{L}_{\text{p-div}}$, is introduced, defined as

$$\mathcal{L}_{\text{p-div}} = \left|\left|\frac{\partial\Delta B_x}{\partial x} + \frac{\partial\Delta B_y}{\partial y}\right|\right|_2^2 \tag{5}$$

where the spatial derivatives are approximated on the discrete 5×5 sensor grid using central finite differences.

While $\mathcal{L}_{\text{p-div}}$ is not a strict 3-D Maxwell constraint, it serves as an effective spatial regularizer. From a signal processing perspective, the magnetic field generated by a physical dipole

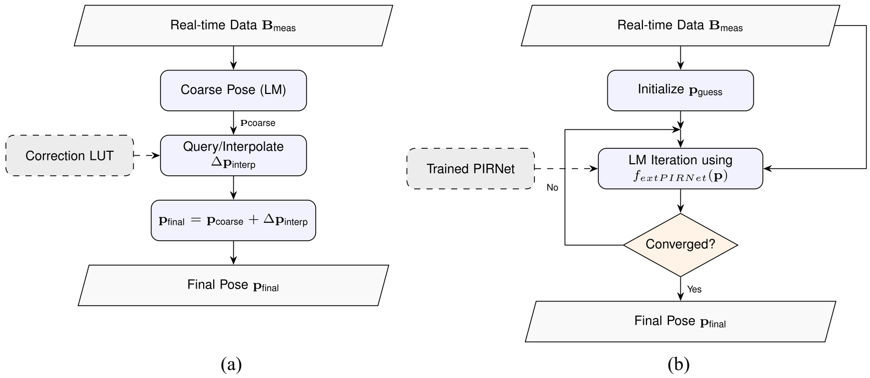


Fig. 2. Operational flowcharts of the online pose estimation paradigms. (a) PIRNet-LUT paradigm utilizing a LUT for rapid compensation. (b) PIRNet-in-the-loop paradigm integrating the neural network within the iterative optimization solver.

is inherently smooth and continuous across the sensor plane. Data-driven neural networks, however, are prone to overfitting sparse noisy data, often producing physically implausible high-frequency spatial jitter in the residual field $\Delta\mathbf{B}$. By minimizing the in-plane divergence, $\mathcal{L}_{\text{p-div}}$ effectively acts as a spatial low-pass filter. It penalizes discontinuous sharp gradient variations along the X- and Y-axes, thereby compelling the network to learn a smooth generalized residual manifold that aligns with the macroscopic physical characteristics of magnetic fields. Alternative formulations exist for enforcing the divergence-free condition. In particular, learning a magnetic vector potential $\mathbf{A}$ and computing the field as $\mathbf{B} = \nabla \times \mathbf{A}$ guarantees $\nabla \cdot \mathbf{B} = 0$ by construction, as the divergence of a curl is identically zero [16]. However, this formulation requires the network to output a 3-D volumetric vector potential field and compute its spatial curl, which, in turn, demands dense 3-D spatial sampling. This requirement is incompatible with the hardware configuration, where all measurements are confined to a single $Z = 0$ planar array. Computing $\frac{\partial A_x}{\partial z}$ or $\frac{\partial A_y}{\partial z}$ from a 2-D sensor grid is as ill-posed as estimating $\frac{\partial \Delta B_z}{\partial z}$ directly. The penalty-based formulation is, therefore, a deliberate engineering tradeoff: it sacrifices strict mathematical enforcement for compatibility with planar sensor hardware, while retaining the physical prior as an effective regularizer.

Subsequent ablation studies demonstrate that this relaxed regularization plays a critical role in suppressing noise and preventing model divergence in real-world scenarios.

C. Paradigm 1: PIRNet-LUT

As illustrated in Fig. 2(a), the first paradigm utilizes the PIRNet offline to generate a residual correction LUT for online inference.

1) LUT Generation: The PIRNet serves as a simulator to quantify intrinsic model discrepancies between the analytical dipole theory and the measured manifold. A grid of M poses $\{\mathbf{p}_j\}_{j=1}^{M}$ is defined across the workspace, constrained by the sensor array's effective range. For each grid pose, the residual error $\Delta\mathbf{p}_j = \mathbf{p}_j - \mathbf{p}_{\text{lm},j}$ is computed, where $\mathbf{p}_{\text{lm},j}$ denotes the estimate from the uncorrected LM algorithm. The LUT stores the error mappings $\{(\mathbf{p}_{\text{lm},j}, \Delta\mathbf{p}_j)\}_{j=1}^{M}$.

2) Online Pose Update: During inference, a new measurement $\mathbf{B}_{\text{meas}}$ yields an initial estimate $\mathbf{p}_{\text{init}}$ consisting of translation $\mathbf{t}_{\text{init}} \in \mathbb{R}^3$ and unit quaternion $\mathbf{q}_{\text{init}} \in \mathbb{H}$. Adjacent error vectors are retrieved and interpolated via trilinear and spherical linear interpolation to determine corrections $\Delta\mathbf{t}_{\text{interp}}$ and $\Delta\mathbf{q}_{\text{interp}}$. The final pose is updated as

$$\mathbf{t}_{\text{final}} = \mathbf{t}_{\text{init}} + \Delta\mathbf{t}_{\text{interp}}, \quad \mathbf{q}_{\text{final}} = \Delta\mathbf{q}_{\text{interp}} \otimes \mathbf{q}_{\text{init}} \tag{6}$$

where $\otimes$ denotes quaternion multiplication to ensure that $\mathbf{q}_{\text{final}}$ remains on the $SO(3)$ manifold. Specifically, the LUT stores orientation corrections as unit quaternions. During online interpolation, the corrections at neighboring grid vertices are combined via Spherical Linear Interpolation (Slerp), which preserves the unit-norm constraint and produces geodesic paths on S^3. The resulting interpolated correction $\Delta\mathbf{q}_{\text{interp}}$ is then composed with the initial estimate via left-multiplication (quaternion product), which corresponds to a rotation composition on the $SO(3)$ manifold. This multiplicative update avoids the singularities and metric distortions inherent in additive Euler angle corrections.

3) Complexity Analysis: The LUT is organized as a 5-D regular grid spanning the operational workspace. The translational axes are discretized as $x, y \in [-80, 80]$ mm at 5-mm intervals (33 points each) and $z \in [60, 140]$ mm at 5-mm intervals (17 points). The orientation, parameterized by polar and azimuthal angles (θ, φ) on S^2, is discretized at 10° intervals, yielding 19 polar and 36 azimuthal points. The total grid size is $33 \times 33 \times 17 \times 19 \times 36 = 12.66 \times 10^6$ entries. Each entry stores a 3-D translational correction and a unit quaternion correction (28 bytes), resulting in a memory footprint of 338 MB, which is compatible with the RAM capacity of contemporary computing platforms including compact x86-based industrial PCs. The $\mathcal{O}(1)$ table query and interpolation complexity enables an execution time of 37.5 ms per frame on Platform C (see Section III-A).

D. Paradigm 2: PIRNet-in-the-Loop

As illustrated in Fig. 2(b), the second paradigm integrates the PIRNet architecture as a differentiable physics layer directly within the iterative optimization process.

The Levenberg–Marquardt (LM) algorithm is selected as the iterative solver because it subsumes the Gauss–Newton method as a special case (when the damping parameter approaches zero) and exploits the analytical Jacobian for locally quadratic convergence, which is categorically more efficient than first-order alternatives for low-dimensional pose estimation with overdetermined residual systems.

1) Differentiable Physics Formulation: The trained PIRNet is defined as a function $f_{\text{PIRNet}}(\mathbf{p})$ mapping a pose vector $\mathbf{p}$ to a corrected magnetic field prediction

$$\mathbf{B}_{\text{corr}}(\mathbf{p}) = \mathbf{B}_{\text{sim}}(\mathbf{p}) + \Delta\mathbf{B}(\mathbf{p}) \tag{7}$$

where $\mathbf{B}_{\text{sim}}(\mathbf{p})$ is the analytical dipole field and $\Delta\mathbf{B}(\mathbf{p})$ is the learned residual.

2) End-to-End Iterative Solver: Online pose estimation is formulated as a nonlinear least-squares problem

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} ||f_{\text{PIRNet}}(\mathbf{p}) - \mathbf{B}_{\text{meas}}||^2. \tag{8}$$

In each optimization iteration, a forward pass through the PIRNet computes the predicted field and the analytical Jacobian $\mathbf{J} = \partial f_{\text{PIRNet}}/\partial \mathbf{p}$ utilizing automatic differentiation. The Jacobian incorporates both the analytical gradient of the dipole model and the learned gradient of the residual field, allowing the LM algorithm to converge on a smoothed physics-informed loss manifold. This paradigm eliminates discretization errors inherent in LUTs and leverages GPU acceleration to achieve a latency of 15.0 ms. The online computational cost of PIRNet-in-the-Loop scales with the number of LM iterations and the cost of one forward-and-Jacobian evaluation of $f_{\text{PIRNet}}(\mathbf{p})$. In contrast to the $\mathcal{O}(1)$ query complexity of PIRNet-LUT after offline table construction, this paradigm trades higher online computation for continuous optimization on the corrected physics manifold.

III. Experimental Setup and Validation

To evaluate the proposed framework, a progressive validation protocol was established, comprising in vitro baseline tracking, an ex vivo signal propagation study through biological tissue, and an in vivo feasibility demonstration. This section details the hardware platforms, experimental scenarios, dataset construction, and evaluation metrics.

A. Hardware Platform

The core hardware comprises a custom magnetic tracking system and independent ground-truth tracking references, depicted in Fig. 3.

1) Magnetic Tracking System: The sensor array is a 5×5 planar grid of triaxial magnetometers utilizing QMC5883L sensors from QST Corp., acquired via a GD32F470ZGT6 microcontroller. Two magnetic targets were fabricated. A standard annular magnet, grade N45 with a 10-mmouter diameter, 5-mm inner diameter, and 10-mm length, is utilized for in vitro and ex vivo characterizations. A miniaturized cylindrical capsule, containing four stacked $\varnothing 3 \times 1$ mm N45 magnets within a biocompatible shell, is utilized for the in vivo animal study.

2) Ground-Truth Acquisition: For in vitro and ex vivo tests, a Mars 2H multicamera optical NOKOV motion capture system provides the baseline 6-degree-of-freedom pose data with a certified accuracy of ± 0.15 mm at 380 frames per second. An Elite Robots EC63 collaborative arm executes predefined trajectories. For the *in-vivo* animal study, optical tracking is structurally obstructed; therefore, the endoscopic video feed and calibrated insertion depth markers are strictly utilized to assess macroscopic trajectory consistency.

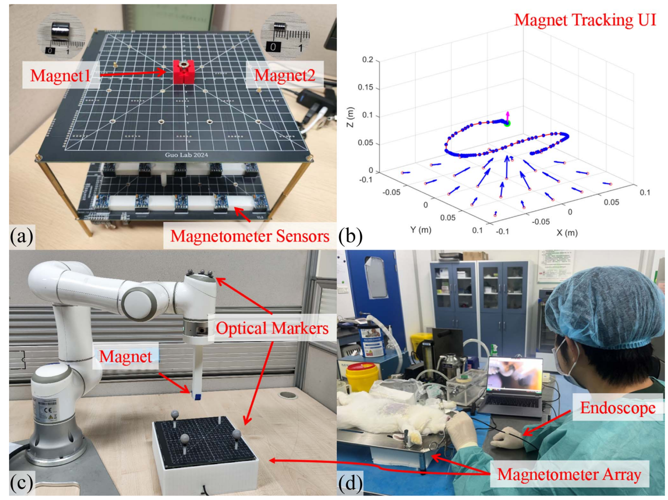


Fig. 3. Experimental platforms. (a) Magnetometer sensor array and magnetic targets. (b) Magnetic tracking user interface, where blue arrows indicate the measured magnetic field vectors. (c) In vitro experimental setup with the optical ground-truth system. (d) In vivo animal study overview.

B. Experimental Scenarios

The validation is structured across four distinct scenarios to isolate specific performance characteristics.

1) In Vitro Baseline: It is conducted in an interference-free environment to establish the absolute tracking accuracy and temporal stability of the proposed paradigms against the uncorrected LM baseline.

2) Ex Vivo Tissue Penetration: It is designed to quantify localization accuracy degradation when the magnetic signal propagates through real biological media. Two fresh porcine tissue blocks, each with a thickness of 30–40 mm and matching human gastrointestinal dielectric and magnetic permeability properties, were positioned directly above the sensor array, while the robotic arm drove the magnet along predefined 3-D trajectories above the tissue.

3) In Vivo Feasibility: It is conducted in a live rabbit model to evaluate system robustness and trajectory consistency under unmodeled dynamic physiological conditions and severe signal attenuation.

4) Dynamic Anti-Interference Compatibility: This scenario involves the introduction of a surgical instrument composed of soft magnetic material into the tracking volume. The objective is to evaluate the capacity of the proposed algorithm to function as a modular observation engine when integrated into a downstream filtering pipeline, such as a system utilizing blind source separation.

C. Datasets and Training Strategy

The dataset construction isolates physical model training from empirical noise adaptation.

1) Prior Model-Based Simulation Data: A total of 58 320 simulated samples were generated on a dense spatial grid spanning [−80, 80] mm in the x- and y-axes and [60, 140] mm in the z-axis. The magnetic dipole model was used to compute the ideal field for each sample, followed by the addition of synthetic Gaussian noise to approximate sensor characteristics.

2) Real-World Measurement Data: A set of 288 unique spatial configurations was collected. At each fixed configuration, 50 temporal measurements were recorded, yielding 14 400 total samples. This temporal resampling isolates the statistical distribution of dynamic sensor noise, ensuring that the spatial calibration burden is explicitly limited to the 288 physical configurations.

3) Training Strategy: To systematically evaluate the framework's data efficiency and its capacity to bridge the sim-to-real gap, a two-phase hybrid training protocol was applied. The network is first pretrained on the simulated data corpus to learn the underlying magnetic dipole physics. It is subsequently fine-tuned on progressively larger subsets of the real-world data, specifically 0%, 5%, 20%, 50%, and 100% of the 288 spatial configurations. This protocol explicitly quantifies the minimal real-world calibration configurations required to adapt the physically grounded prior to the intrinsic sensor noise characteristics and static model-plant mismatches.

D. Evaluation Metrics

Localization performance is evaluated using pointwise position and orientation errors. Let the ground-truth pose be represented by a position vector $\mathbf{t}_{\text{gt}} \in \mathbb{R}^3$ and an orientation unit vector $\mathbf{v}_{\text{gt}} \in \mathbb{R}^3$, where $||\mathbf{v}_{\text{gt}}||_2 = 1$. The corresponding estimated pose is denoted as $\mathbf{t}_{\text{est}}$ and $\mathbf{v}_{\text{est}}$.

1) *Position error (E_{pos})* measures the Euclidean distance between the estimated and true positions

$$E_{\text{pos}} = ||\mathbf{t}_{\text{est}} - \mathbf{t}_{\text{gt}}||_2 \tag{9}$$

2) *Orientation error (E_{ori})* measures the angular difference in degrees between the estimated and true orientation vectors

$$E_{\text{ori}} = \arccos(\mathbf{v}_{\text{est}} \cdot \mathbf{v}_{\text{gt}}) \cdot \frac{180}{\pi}. \tag{10}$$

To ensure statistical consistency, all aggregated localization performances across the datasets are exclusively reported as the Mean ± Standard Deviation (Std Dev) of these pointwise errors.

For the in vivo feasibility study, where point-to-point submillimeter ground truth is unavailable, two macroscopic metrics are employed to evaluate structural integrity.

1) *Track length:* It evaluates the presence of high-frequency spatial jitter by comparing the integrated length of the estimated trajectory against the known endoscopic insertion depth.

2) *Fréchet distance:* It quantifies the geometric shape similarity between two continuous trajectory curves to evaluate structural consistency.

TABLE II
PERFORMANCE BENCHMARK

Method	Pos. Error (mm)	Ori. Error (°)	Latency (ms)
Literature Benchmarks (Reported values)			
Guo et al. [22]	0.90 ± 0.75	1.51 ± 0.87	79.3
Su et al. [18]	1.87 ± 1.14	1.89 ± 0.81	2.08
Xie et al. [11]	1.54 ± 1.03	2.24 ± 1.84	**0.90**
Cai et al. [6]	0.80 ± 0.15	**0.12 ± 0.07**	7.00
Internal Benchmarks (Our dataset)			
Baseline (LM on Dipole)	1.72 ± 1.04	5.57 ± 1.60	36.0
Data-Driven ResNet (100%)	1.45 ± 0.38	1.95 ± 0.65	3.30
Ours (PIRNet-LUT)	1.01 ± 0.35	1.38 ± 0.51	37.5
Ours (PIRNet-in-the-Loop)	**0.69 ± 0.29**	0.70 ± 0.37	15.0

Literature benchmarks report values from the respective publications under heterogeneous hardware configurations and are provided as contextual references. Internal benchmarks are evaluated on the present dataset under identical conditions. Bold values denote the best results among the compared methods for each metric.

E. Implementation Details

The PIRNet was implemented using the PyTorch framework. The PIRNet-LUT and PIRNet-in-the-Loop operational paradigms, along with data processing, were executed in MATLAB R2023a. The Python-based neural network inference was called via MATLAB environment interfaces.

Neural network training was performed on a server configured with an NVIDIA RTX 4090 GPU and an Intel Platinum 8368 CPU. To assess deployment feasibility under varying computational constraints, inference latency was benchmarked across three hardware platforms: Platform A featuring an Intel Core i7-12700 CPU and an NVIDIA GeForce GTX 1660Ti GPU; Platform B featuring an Intel Core i5-7300HQ CPU and an NVIDIA GeForce GTX 1050 GPU; and Platform C featuring an AMD Ryzen 7 8845HS CPU utilizing CPU-only inference.

IV. Results and Analysis

This section evaluates the proposed framework. The core performance and architectural contributions of the PIRNet model are analyzed, followed by a comparison of the operational paradigms across hardware platforms and an investigation of data efficiency. Dynamic tracking capabilities are then validated, progressing from an in vitro baseline to an ex vivo signal propagation study through biological tissue, and concluding with an in vivo feasibility demonstration.

A. Core Performance and Ablation

The primary function of the PIRNet framework is to learn a physically plausible residual correction bridging the gap between the idealized magnetic dipole model and real-world sensor measurements. A visual representation of this process is provided in Fig. 4.

As detailed in Table II, the poses estimated via the PIRNet paradigms demonstrate improved position and orientation accuracy compared to the uncorrected LM baseline. This indicates

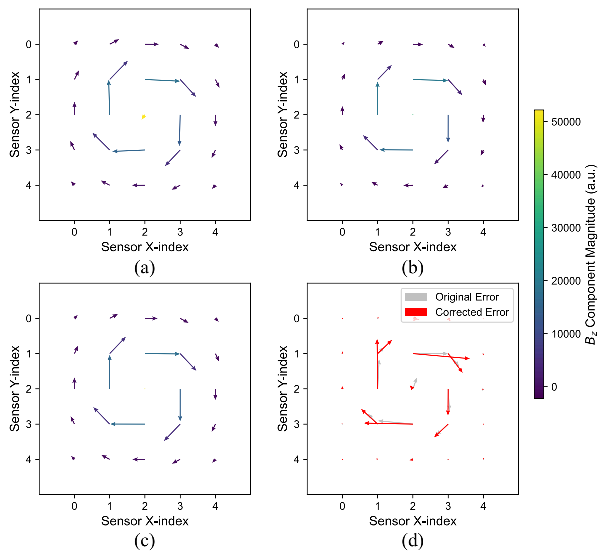


Fig. 4. Magnetic field correction visualization for a sample pose. (a) Ideal field from the dipole model ($\mathbf{B}_{\text{sim}}$). (b) Measured field from real sensors ($\mathbf{B}_{\text{real}}$). (c) PIRNet-corrected field ($\mathbf{B}_{\text{corr}}$). (d) Error vector fields comparing $\mathbf{B}_{\text{real}} - \mathbf{B}_{\text{sim}}$ (gray) and $\mathbf{B}_{\text{real}} - \mathbf{B}_{\text{corr}}$ (red, scaled by 2x for visibility).

TABLE III
ABLATION STUDY OF PIRNET COMPONENTS

Model Configuration	Position Error (mm)	Orientation Error (°)
Ours (Full PIRNet)	**0.69 ± 0.29**	**0.70 ± 0.37**
w/o Physics Branch	1.78 ± 0.89	1.95 ± 1.02
w/o Gated Fusion	1.65 ± 0.75	1.82 ± 0.93
w/o $\mathcal{L}_{\text{p-div}}$	2.41 ± 1.33	2.88 ± 1.57

Bold values denote the best results among the compared methods for each metric.

that the dual-branch architecture disentangles and models systematic model-plant mismatches and localized nonlinear sensor distortions.

To evaluate the contributions of the architectural components, ablation studies were conducted on the static test set. As summarized in Table III, removing individual components degrades localization performance. Notably, training without the in-plane pseudodivergence regularization (w/o $\mathcal{L}_{\text{p-div}}$) results in the largest performance drop. This underscores that enforcing a relaxed spatial constraint acts as a spatial low-pass filter, preventing the network from fitting high-frequency sensor noise and maintaining the physical plausibility of the learned residual manifold.

B. Comparison of Operational Paradigms

PIRNet-LUT and PIRNet-in-the-Loop are compared in terms of accuracy and computational latency across three hardware platforms. The results, presented in Fig. 5 and detailed in Table IV, indicate a hardware-dependent tradeoff, which is further discussed in Section V-A.

C. Data Efficiency Analysis

To evaluate few-shot capabilities under identical conditions, a comparative study was conducted using 288 real-world spatial configurations. A data-driven baseline (*Data-Driven ResNet*) was implemented as a controlled ablation of the physics-informed components. This baseline matches the convolutional encoder capacity of PIRNet (identical layer count, channel dimensions, and residual block structure) but removes three physics-informed elements simultaneously: the pose-dependent physics encoder branch, the simulation-based pre-training phase, and the $\mathcal{L}_{\text{p-div}}$ regularization. This design isolates the aggregate contribution of all physics-informed components under matched network capacity, providing a stricter test than comparing against a weaker MLP architecture. Both models were trained or fine-tuned using varying fractions of the real-world dataset: 0%, 5% (14 configurations), 20%, 50%, and 100%. To ensure a controlled comparison, both models utilized the identical temporal resampling strategy of 50 repeated measurements per configuration across all data splits.

The learning curves in Fig. 6 characterize the convergence behavior of position error [see Fig. 6(a)] and orientation error [see Fig. 6(b)] relative to the calibration data volume. In the low-data regime involving less than 5% of the configurations, the data-driven ResNet baseline exhibits position errors exceeding 4 mm and orientation errors exceeding 10°. This baseline requires the full dataset of 288 configurations to reach a mean position error of 1.45 mm. The PIRNet exhibits a different convergence profile. At 0% real-world data, the model yields a mean position error of 1.78 mm and a mean orientation error of 5.85°, which are comparable to the analytical LM baseline (1.72 mm/5.57°). This initial performance is attributed to the analytical dipole prior, while the pretrained residual branch

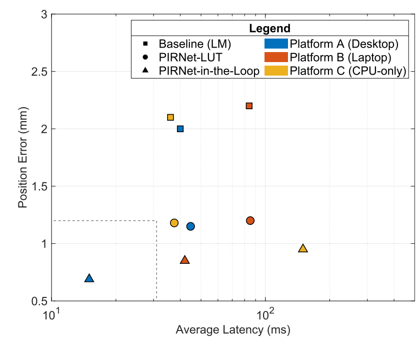
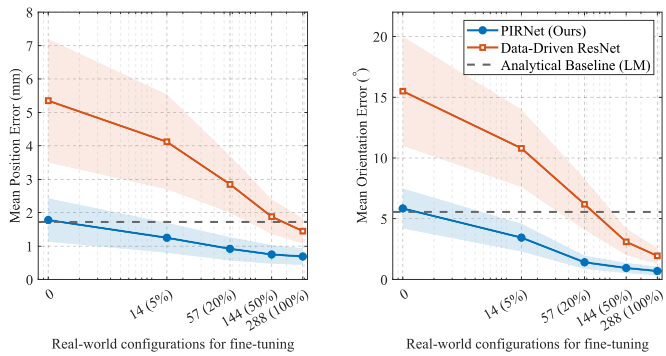


Fig. 5. Accuracy versus latency evaluated across three hardware platforms.

TABLE IV
COMPUTATIONAL LATENCY BENCHMARK ON DIFFERENT PLATFORMS

Paradigm	Platform	Aver. Latency (ms)	Real-Time?
Baseline (LM)	Platform A	40.0	Yes (25 FPS)
	Platform B	84.0	Marginally (12 FPS)
	Platform C	36.0	Yes (28 FPS)
PIRNet-LUT	Platform A	44.7	Yes (22 FPS)
	Platform B	85.0	Marginally (12 FPS)
	Platform C	37.5	Yes (27 FPS)
PIRNet-in-the-Loop	Platform A	**15.0**	**Yes (67 FPS)**
	Platform B	42.0	Yes (24 FPS)
	Platform C	150.0	No (7 FPS)

Bold values denote the best results among the compared methods for each metric.



(a) (b)

Fig. 6. Data efficiency analysis under identical experimental conditions. (a) Position error and (b) orientation error as functions of the proportion of real-world calibration configurations. PIRNet (blue) converges with minimal data, whereas the data-driven ResNet baseline (red) requires the full dataset to approach comparable accuracy.

outputs corrections derived from simulated noise distributions. As real-world configurations are introduced, the position error decreases to 0.92 mm, and the orientation error decreases to 1.42° at a 20% data fraction (57 configurations). These results indicate that the physics-informed architecture reaches the accuracy threshold of the fully trained data-driven baseline using less than one-fifth of the real-world calibration configurations for both position and orientation metrics.

D. Dynamic Tracking Validation

Dynamic tracking performance was evaluated through in vitro baseline tracking, an ex vivo signal propagation study through biological tissue, and an in vivo feasibility demonstration.

1) In Vitro Baseline: Baseline dynamic tracking accuracy was evaluated in an interference-free in vitro environment using the 10-mm magnet. As illustrated in Fig. 7, both PIRNet paradigms follow the optical ground truth. The 2-D error analysis (see Fig. 8) demonstrates that the proposed methods suppress the spatial jitter present in the uncorrected LM baseline. A preliminary static evaluation of the 3-mm miniaturized capsule yielded a baseline position error of 1.64 $\pm$ 0.97 mm, which establishes the reference for subsequent low-SNR tests.

2) Ex Vivo Tissue Tracking: To evaluate the framework under anatomical constraints, an ex vivo porcine tissue study was conducted [see Fig. 9(a)]. Two porcine tissue blocks, each 30–40 mm thick, were positioned above the sensor array, and the magnetic target was encapsulated within porcine hepatic tissue. To isolate the electromagnetic influence of the biological media, the experimental volume and sensor-to-target distances were maintained consistent with the in vitro trials, with zero mechanical contact between the encapsulated target and the stationary tissue blocks.

As biological tissues possess a relative magnetic permeability (μ_r) approximately equal to unity, they do not induce measurable attenuation or distortion of low-frequency magnetic fields. The results, summarized in Fig. 9(b), confirm this physical principle. Quantitatively, the PIRNet-in-the-Loop paradigm yielded a mean position error of 0.74 $\pm$ 0.32 mm and an orientation error of 0.76 $\pm$ 0.41°. The PIRNet-LUT approach achieved 1.08 $\pm$ 0.39 mm and 1.45 $\pm$ 0.58°, while the baseline LM algorithm yielded 1.85 $\pm$ 1.15 mm and 5.82 $\pm$ 1.74°.

TABLE V
QUANTITATIVE ANALYSIS OF TRACKED TRAJECTORIES IN THE IN VIVO EXPERIMENT

Method	Track Len. (cm)	Fréchet Dist. (cm)
Endoscope Depth Marker (GT)	18.5 (Reference)	-
Baseline (LM on Dipole)	20.3	1.54
Ours (PIRNet-LUT)	**18.3**	**0.95**
Ours (PIRNet-in-the-Loop)	**19.0**	**0**

Note: In the absence of an independent ground truth, the Fréchet distance is computed against the PIRNet-in-the-Loop trajectory as a pseudoreference for relative comparison.
Bold values denote the best results among the compared methods for each metric.

The marginal increments in error (less than 0.1 mm) compared to the in vitro results are attributed to stochastic sensor thermal drift and minor variations in the robotic arm's dynamic load during tissue-encapsulated motion, rather than biological interference. The high consistency between the air and tissue experiments provides quantitative evidence that the PIRNet framework maintains its localization accuracy when propagating through biological media, established by a reliable optical ground truth.

3) In Vivo Feasibility: A preliminary in vivo experiment was conducted using a 2.2-kg New Zealand White Rabbit. The miniaturized capsule was attached to an endoscope and withdrawn from the stomach (see Fig. 10).

In the absence of a synchronized submillimeter ground truth in this live setup, macroscopic trajectory consistency metrics are utilized. Fig. 10(a) visualizes the tracked trajectories during three repeated withdrawals. Table V reports Track Length and Fréchet Distance. The uncorrected LM baseline exhibits spatial jitter, overestimating the 18.5-cm reference length to 20.3 cm. The PIRNet paradigms output continuous trajectories that more closely match the actual endoscopic insertion depth. Since no independent submillimeter ground truth is available in the live-animal setting, PIRNet-in-the-Loop is adopted as a *pseudo-reference* for relative trajectory comparison. Under this metric, the PIRNet-LUT approach yields a Fréchet Distance of 0.95 cm. These results suggest that the learned residual manifold and $\mathcal{L}_{\text{p-div}}$ regularization suppress nonphysical spatial jitter under in vivo conditions, supporting the potential for engineering integration into closed-loop systems. Definitive accuracy quantification awaits the availability of a synchronized radiological ground-truth system such as biplane fluoroscopy.

E. Anti-Interference Compatibility

Performance under dynamic ferromagnetic interference was evaluated following the protocol defined in Scenario 4. A downstream filtering architecture, comprising a linear blind source separation module and an adaptive Kalman filter, was employed to reject the induced distortions.

The quantitative comparisons during the active interference interval are presented in Fig. 11. Utilizing the uncorrected LM model without the downstream filter yielded a mean position error of 3.41 $\pm$ 2.78 mm and an orientation error of 15.51 $\pm$ 6.66°. Integrating the LM model as the observation solver within this filter resulted in a mean position error of 3.31 $\pm$ 2.74 mm and

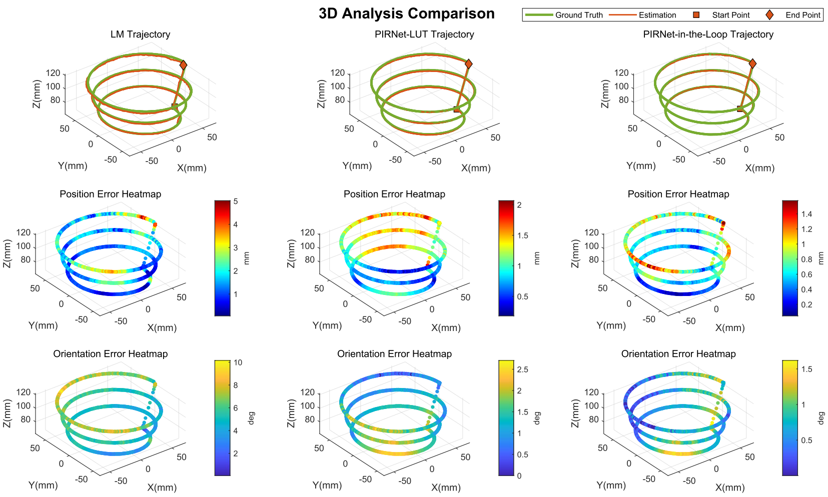


Fig. 7. 3-D analysis and comparison of dynamic tracking results in the in vitro experiment. The top row shows 3-D tracked trajectories. The middle and bottom rows visualize pointwise position and orientation errors as heatmaps.

Fig. 8. Detailed 2-D error analysis of the in vitro dynamic tracking experiment over time (top) and as a function of magnet height (bottom).

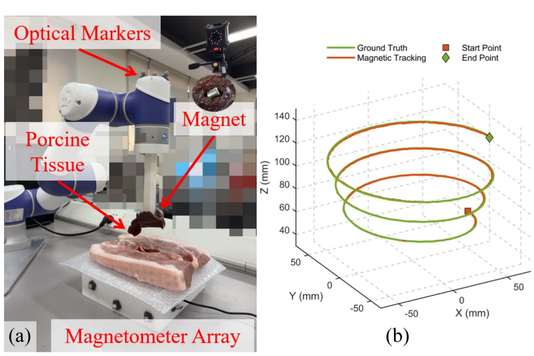


Fig. 9. Quantitative ex vivo dynamic tracking performance through porcine tissue. (a) Experimental setup. (b) 3-D trajectory comparison highlighting the alignment of the proposed PIRNet methods with the optical ground truth.

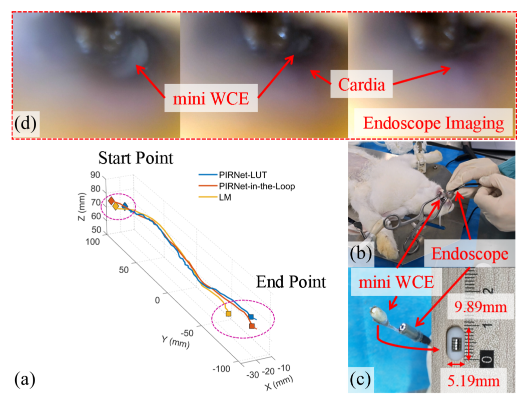


Fig. 10. In vivo animal study results. (a) 3-D tracked trajectories of three repeated withdrawal motions of the mini-capsule. (b) Experimental site. (c) Endoscope and miniaturized WCE. (d) Sequential endoscopic video frames.

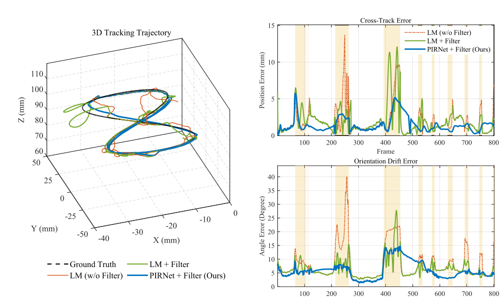


Fig. 11. Position and orientation errors during dynamic tracking under active scissor interference. The shaded region indicates the soft-magnetic distortion interval. The curves compare the uncorrected LM baseline (w/o Filter), the LM model integrated with the downstream filter (LM + Filter), and the integrated PIRNet-in-the-Loop observation model (PIRNet + Filter).

an orientation error of $9.41 \pm 4.65^\circ$. In contrast, the integrated system utilizing the PIRNet-in-the-Loop paradigm maintained a mean position error of 2.03 ± 1.44 mm and an orientation error of $7.86 \pm 3.49^\circ$.

Within the evaluated integrated pipeline, these outcomes indicate that the PIRNet paradigm provides bounded residuals that allow the downstream filter to isolate induced distortions, supporting compatibility with industrial filtering architectures.

Isolating the individual contribution of each pipeline stage requires further controlled ablation.

V. Discussion

The experimental results presented in Section IV support the effectiveness of the PIRNet framework under the evaluated conditions. This section analyzes the performance tradeoffs, discusses implications for model generalization, and defines the operational boundaries of the current study.

A. Performance Tradeoffs

The operational paradigms indicate hardware-dependent tradeoffs between localization fidelity and computational latency. Table II compares the PIRNet framework with contemporary literature and internal baselines.

The literature benchmarks in Table II are reproduced from the respective publications and correspond to different sensor array geometries, magnet specifications, and workspace ranges. As a result, a controlled head-to-head evaluation would require reimplementing each system on identical hardware, which is infeasible given the proprietary sensor configurations involved. The cross-study values are, therefore, provided solely to contextualize the performance envelope of the proposed framework, rather than to establish strict superiority claims. The internal benchmarks, evaluated on the present dataset under identical conditions, constitute the primary basis for the conclusions drawn in this study.

Within the contextual literature values summarized in Table II, the PIRNet-in-the-Loop paradigm operates in the submillimeter accuracy range while maintaining real-time performance on GPU-enabled platforms. While the MobilePosenet proposed by Xie et al. demonstrates the minimum computational latency at 0.90 ms, the localization accuracy remains limited to 1.54 mm. The PIRNet-in-the-Loop solver provides an alternative for applications requiring submillimeter precision. Furthermore, on GPU platforms, the refined loss landscape of the PIRNet reduces the iterations necessary for convergence, maintaining an execution time of 15.0 ms. The PIRNet-LUT paradigm maintains real-time performance on CPU-constrained devices by decoupling network inference from the online optimization process. Compared to the purely data-driven ResNet baseline, the physics-informed paradigms demonstrate higher accuracy while utilizing a fraction of the real-world configurations, which addresses the calibration burden inherent in conventional end-to-end models.

B. Data Efficiency and Model Generalization

Integrating physical priors reduces the data dependence inherent in conventional deep learning architectures. The few-shot analysis demonstrates that the PIRNet converges using less than 20% of the calibration dataset required by the purely data-driven ResNet baseline. The analytical dipole model and the in-plane pseudodivergence penalty function as inductive biases, suppressing the fitting of unmodeled spatial noise. This reduction in data requirements facilitates rapid system calibration.

The framework further demonstrates generalization capability across varying magnet geometries. Validated on a standard annular magnet and a miniaturized multimagnet capsule, the PIRNet adapts to specific model-plant mismatches via sparse fine-tuning without architectural modifications. This indicates that the learned residual manifold applies to diverse magnetic profiles utilized in medical robotic applications.

The current ablation study isolates the contributions of the physics-informed components under a fixed convolutional backbone. Evaluation of alternative backbone architectures, such as vision transformers or equivariant networks, constitutes an orthogonal research direction that may yield further improvements and is reserved for future investigation.

C. Limitations and Future Work

The operational boundaries of the current study are defined as follows.

First, the performance benchmarks reported against prior literature in Table II are based on values reported in the respective publications, as the heterogeneous sensor hardware and magnet specifications across studies preclude a controlled head-to-head evaluation on a shared dataset. Consequently, the cross-study comparison serves as a contextual reference rather than a controlled experimental benchmark. The development of a community-wide standardized evaluation benchmark—employing a shared sensor array geometry, magnet specification, and reference dataset—would be valuable for enabling fair cross-method comparisons in future work.

Second, the in-plane pseudodivergence regularization $\mathcal{L}_{\text{p-div}}$ constitutes a relaxed physical constraint rather than a strict enforcement of Maxwell's divergence-free condition. Alternative formulations, such as learning a magnetic vector potential $\mathbf{A}$ and computing $\mathbf{B} = \nabla \times \mathbf{A}$, guarantee a divergence-free field by construction. However, such approaches require volumetric 3-D spatial information that is unavailable from a single planar sensor array. The proposed penalty-based approach trades strict physical enforcement for hardware compatibility, and its empirical efficacy is supported by the ablation results in Table III.

Third, training and testing were conducted on the same sensor array platform within a controlled laboratory environment. The robustness of the learned residual manifold under out-of-distribution conditions remains unquantified along several axes.

1) Cross-array generalization—transferring a trained model to a sensor array with different grid spacing or sensor model—has not been tested.
2) A systematic characterization of localization accuracy as a function of sensor-to-target distance, which directly governs the signal-to-noise ratio and defines the operational workspace boundaries, is absent.
3) Quantified temperature-sweep experiments spanning the clinically relevant range of 20 °C–40 °C have not been conducted to isolate thermal drift effects from other noise sources.

These constitute prerequisites for clinical deployment and are prioritized directions for future investigation. The development of a community-wide standardized evaluation benchmark—employing a shared sensor array geometry, magnet specification, and reference dataset—would further enable fair cross-method comparisons. To facilitate reproducibility and future benchmarking, the sensor array hardware schematics and the baseline LM solver implementation are publicly available online.[1]

Fourth, the in vivo animal study lacks a synchronized submillimeter radiological ground-truth system such as biplane fluoroscopy. The in vivo evaluation is, therefore, restricted to macroscopic trajectory consistency analysis and does not constitute quantitative clinical validation.

Finally, extending this framework to the simultaneous tracking of multiple active hard-magnetic sources from a single planar array remains a mathematically underdetermined problem. The superposition of multiple nonlinear dipole fields confounds the residual manifold learned by the current architecture, necessitating future research into permutation-invariant network structures or temporal source-decoupling strategies.

VI. Conclusion

This study addressed the model-plant mismatch in magnetic localization for wireless capsule endoscopy through a physics-informed deep learning framework based on the PIRNet architecture. Two operational paradigms are formulated and evaluated: the real-time PIRNet-LUT approach and the differentiable PIRNet-in-the-Loop solver.

Experiments indicated that the framework yields a position error of 0.69 mm and adapts to different magnet geometries. By integrating physical priors, the framework reduced the dependence on real-world training data, requiring less than one-fifth of the configurations necessary for purely data-driven baselines.

System feasibility was evaluated through a progressive protocol. An ex vivo signal propagation study through biological tissue provided quantitative evidence that tracking accuracy is maintained through biological media, compatibility tests under dynamic ferromagnetic interference supported integration with downstream filtering pipelines, and a preliminary in vivo animal study supported the qualitative applicability of the framework in a living-subject setting.

References

[1] Y. Sun et al., "Magnetically driven capsules with multimodal response and multifunctionality for biomedical applications," *Nature Commun.*, vol. 15, no. 1, Feb. 2024, Art. no. 1839.

[2] Y. Yan et al., "5-DOF microcoil positioning system utilizing single-axis electromagnetic transmitter," *IEEE/ASME Trans. Mechatron.*, vol. 30, no. 6, pp. 5213–5224, Dec. 2025, doi: 10.1109/TMECH.2025.3554246.

[3] L. Zheng, S. Guo, and M. Kawanishi, "Multimodular capsule robot system with drug release for intestinal treatment," *IEEE Sensors J.*, vol. 23, no. 16, pp. 18568–18578, Aug. 2023.

[4] M. Mascarenhas et al., "The future of minimally invasive capsule panendoscopy: Robotic precision, wireless imaging and AI-Driven insights," *Cancers*, vol. 15, no. 24, Dec. 2023, Art. no. 5861.

[5] M. A. Ali, N. Tom, F. N. Alsunaydih, and M. R. Yuce, "Recent advancements in localization technologies for wireless capsule endoscopy: A technical review," *Sensors*, vol. 25, no. 1, Jan. 2025, Art. no. 253.

[1][Online]. Available: https://github.com/ShenMiaozhang/sensor-array-hardware-lm-baseline

[6] Z. Cai, M. Wang, H. Wang, J. Han, and Y. Qin, "High-precision magnetic tracking with individual magnet identification for medical applications," *IEEE/ASME Trans. Mechatron.*, early access, 20 Feb. 2026, doi: 10.1109/TMECH.2026.3660131.

[7] L. Zheng, S. Guo, and M. Kawanishi, "Magnetically controlled multifunctional capsule robot for dual-drug delivery," *IEEE Syst. J.*, vol. 16, no. 4, pp. 6413–6424, Dec. 2022.

[8] Q. Cao et al., "Robotic wireless capsule endoscopy: Recent advances and upcoming technologies," *Nature Commun.*, vol. 15, no. 1, May 2024, Art. no. 4597.

[9] B. Lv, Y. Qin, H. Dai, and J. Han, "A differential signal and extended-Kalman-filter-based anti-interference magnetic tracking method for surgical scenes," *IEEE Trans. Ind. Electron.*, vol. 71, no. 11, pp. 15140–15150, Nov. 2024.

[10] Z. Wang, S. Guo, J. Guo, Q. Fu, L. Zheng, and T. Tamiya, "Selective motion control of a novel magnetic-driven minirobot with targeted drug sustained-release function," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 1, pp. 336–347, Feb. 2022.

[11] W. Xie, Y. Zhang, J. Shan, H. Sun, J. Tan, and S. S. Cheng, "Theoretical data-driven MobilePosenet: Lightweight neural network for accurate calibration-free 5-DOF magnet localization," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 2536511.

[12] W. Chen, J. Sui, and C. Wang, "Magnetically actuated capsule robots: A review," *IEEE Access*, vol. 10, pp. 88398–88420, 2022.

[13] S. Su, S. Yuan, Z. Li, Y. Ma, M. Ma, and H. Ren, "A wearable, reconfigurable, and modular magnetic tracking system for wireless capsule robots," *IEEE Trans. Ind. Informat.*, vol. 20, no. 12, pp. 13600–13611, Dec. 2024.

[14] J. Luo, Q. Huang, H. Dai, S. Wang, and Y. Chen, "Adaptively adjusted EKF-based magnet tracking method for fast-moving object," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 7502309.

[15] Y. Fu and Y. Guo, "Guideline on initial point finding and search bounds setting for biomedical magnetic localization," *IEEE Sensors J.*, vol. 22, no. 16, pp. 16492–16504, Aug. 2022.

[16] P. Ernst et al., "A physics-informed neural network for the calibration of electromagnetic navigation systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2025, pp. 5743–5749.

[17] S. Song, H. Ge, J. Wang, and M. Q.-H. Meng, "Real-time multi-object magnetic tracking for multi-arm continuum robots," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 4008309.

[18] S. Su, S. Yuan, M. Xu, H. Gao, X. Yang, and H. Ren, "AMagPoseNet: Real-time six-DoF magnet pose estimation by dual-domain few-shot learning from prior model," *IEEE Trans. Ind. Informat.*, vol. 19, no. 9, pp. 9722–9732, Sep. 2023.

[19] Q. Fu, D. Zhao, L. Shao, and S. Zhang, "Magnetic localization algorithm of capsule robot based on BP neural network," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 4002509.

[20] M. Wu, T. Langerak, O. Hilliges, and J. Zarate, "Usynthetic datasing in supervised learning for robust 5-DoF magnetic marker localization," *IEEE Trans. Magn.*, vol. 60, no. 8, Aug. 2024, Art. no. 4000211.

[21] T. Wen, S. Liu, and H. Wang, "Soft-magnet-based position estimation using an approximate extended Kalman filter with a hybrid analytical-neural network model," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 7500113.

[22] P. Guo, H. Dai, Q. Yang, Q. Huang, and H. Yao, "An improved magnetic tracking approach based on ResNet-LM fusion algorithm," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 2520510.

[23] N. Sebkhi et al., "A deep neural network-based permanent magnet localization for tongue tracking," *IEEE Sensors J.*, vol. 19, no. 20, pp. 9324–9331, Oct. 2019.

[24] J. Lu, F. Zheng, S. Zhang, Q. Yang, and C. Hu, "Regional perceptrons ensemble for permanent magnetic localization," *IEEE Sensors J.*, vol. 24, no. 17, pp. 28056–28067, Sep. 2024.

[25] S. Su, H. Lan, H. Gao, H. Dai, and H. Ren, "Enhancing anti-interference of magnetic tracking: A MagRobustNet-based framework with self-supervised anomaly detection and measurements recovery," *IEEE Trans. Ind. Informat.*, vol. 21, no. 1, pp. 781–790, Jan. 2025.

[26] C. Hu, M. Q. Meng, and M. Mandal, "Efficient magnetic localization and orientation technique for capsule endoscopy," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 628–633.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

Miaozhang Shen (Graduate Student Member, IEEE) received the B.S. degree in measurement and control technology and instrumentation from Shenzhen University, Shenzhen, China, in 2022. He is currently working toward the M.S. degree in electronic information with the Southern University of Science and Technology, Shenzhen.

His research interests include magnetic medical robot systems.

Shuxiang Guo (Fellow, IEEE) received the Ph.D. degree in mechano-informatics and systems from Nagoya University, Nagoya, Japan, in 1995.

He is currently a Chair Professor with the Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, China, and also with the Key Laboratory of Convergence System and Healthcare Technology Medical Engineering, Ministry of Industry and Information Technology, Beijing Institute of Technology, Beijing, China. His research interests include medical robot systems, microcatheter systems, and biomimetic underwater robots.

Dr. Guo is a Fellow of the Engineering Academy of Japan and the Editor-in-Chief for *International Journal of Mechatronics and Automation*.

Zixu Wang (Member, IEEE) received the Ph.D. degree in intelligent mechanical systems engineering from Kagawa University, Takamatsu, Japan, in 2021.

He is currently a Postdoctoral Researcher with the Southern University of Science and Technology, Shenzhen, China. His research interests include medical robot systems for minimal invasive surgery and magnetic-driven flexible robot.

Yuyue Yang received the B.S. degree in optoelectronic information science and engineering from Shenzhen University, Shenzhen, China, in 2022. He is currently working toward the M.S. degree in integrated circuit engineering with South China Normal University, Guangzhou, China.

His research interests include computational imaging and optical design.

Chunying Li (Senior Member, IEEE) received the Ph.D. degree in intelligent mechanical systems engineering from the Graduate School of Engineering, Kagawa University, Takamatsu, Japan, in 2023.

He is currently a Research Assistant Professor with the Advanced Institute for Ocean Research, Southern University of Science and Technology, Shenzhen, China. His research interests include bionic underwater robots, control systems, multisensor information fusion, and multirobot collaboration.

Mingchao Ding received the M.S. degree in imaging and nuclear medicine from the School of Medicine, Peking University, Beijing, China, in 2011.

He is currently a Professor, Chief Physician, and Vice President with Aerospace Center Hospital, Beijing. He also directs the National Health Commission Continuing Education Base for Peripheral Vascular Interventional Physicians. His research interests include endovascular and emergency interventional treatment of ischemic, bleeding, and thrombotic vascular diseases.

Bing Wang received the M.S. degree in endocrinology and metabolic diseases from Shandong University, Jinan, China, in 2007.

He is currently the Director of the Interventional Vascular Department, Chief Physician, and Master's Supervisor with Aerospace Center Hospital, Beijing, China. His expertise includes interventional treatment of peripheral vascular diseases, deep vein thrombosis, pulmonary embolism, ischemic vascular diseases, and minimally invasive treatment of varicose veins.