

Residual Neural Network for Precise 6-DoF Capsule Endoscope Localization Using Electromagnetic Induction

Van Luong Tran, Xuan Mai Nguyen, Dong Eui Chang, *Member, IEEE*, Si Hong Hoang, Manh Cuong Hoang

Abstract—Accurate localization of capsule endoscopes is essential for advancing minimally invasive gastrointestinal diagnostics. Traditional magnetic induction-based methods, which rely on magnetic model and nonlinear optimization, often encounter convergence issues due to local minima and sensitivity to initial conditions, limiting accuracy and robustness, especially under dynamic orientations. This paper proposes a deep learning approach to estimate the six degrees-of-freedom pose-position (x , y , z) and orientation (roll, pitch, yaw)—of a capsule endoscope from 9 electromotive force (EMF) measurements. These measurements are generated via an induction method using a 3 transmitting coil system powered by AC currents and a tri-axis receiving coil. The proposed model integrates residual blocks with skip connections, enabling effective learning of complex, nonlinear mappings compared to standard fully connected neural networks. This architecture enhances generalization across sharp orientation changes and improves robustness to noise. Additionally, we introduce a compact, wireless electronic module for real-time EMF signal acquisition and transmission, enabling untethered operation in capsule endoscope system. Experimental results under dynamic conditions demonstrate a mean position error below 2 mm, an orientation error under 4°, and inference times under 1 ms. These findings validate the suitability of the system for real-time, closed-loop magnetic navigation in clinical environments without dependence on external computation.

Index Terms—Neural network, residual block, capsule endoscope localization, electromagnetic induction, wireless capsule endoscope, nonlinear optimization algorithm

I. INTRODUCTION

IN the context of the growing demand for modern and minimally invasive healthcare, wireless capsule endoscopy (WCE) is increasingly developing as a promising future technology for applications such as gastrointestinal disease diagnosis [1]–[4], biopsy [5]–[7], and drug delivery [8]–[10] inside the human body. However, important WCE requirements, such as precise control [11], or the need for doctors to know the exact location of pathological tissues when WCE captures images [12]–[14], require accurate localization of the

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2024-XS-003. (Corresponding authors: Manh Cuong Hoang.)

Van Luong Tran and Dong Eui Chang are with the Control Lab, School of Electrical Engineering, Korea Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: luongtv1512@kaist.ac.kr; dechang@kaist.ac.kr)

Xuan Mai Nguyen is with the School of Electrical Engineering, Korea Institute of Science and Technology, Daejeon 34141, (e-mail: mainx@kaist.ac.kr).

Si Hong Hoang and Manh Cuong Hoang are with the School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi 100000, Vietnam (e-mail: hong.hoangsi@hust.edu.vn, cuong.hoangmanh@hust.edu.vn)

Manuscript received ; revised

capsule. One of the major challenges in localization lies in the limited space available for attaching positioning devices, which makes accurate localization difficult. To address this issue, researchers worldwide have focused on researching and developing localization technologies to track WCE.

Various magnetic localization methods have been proposed for capsule endoscopy, with two main sensing approaches: Hall-effect sensor arrays and electromagnetic induction-based systems. Hall-effect sensor arrays have been widely used to estimate up to 5-DOF. Son *et al.* [15] developed a 5-DOF localization system using a 2D Hall sensor array. Pham and Aziz *et al.* [16] proposed a magnet localization method based on magnetic field sensors. Other works [17]–[20] explored different array configurations to improve accuracy and coverage. More recently, Hongxi *et al.* [21] introduced reconfigurable magnetic sensor arrays to reduce both position and orientation errors during tracking. In contrast, electromagnetic induction-based localization offers the potential to estimate the full 6-DOF pose of the capsule. Islam *et al.* [22], [23] proposed a system consisting of three ring-shaped transmitting (Tx) coils and three orthogonal receiving (Rx) coils embedded in the capsule. Liu *et al.* [24] utilized three non-orthogonal Tx coils and three nonorthogonal Rx to achieve 6-DOF localization using a nonlinear optimization algorithm. Song *et al.* [25] developed a method using a three-axis sensing Rx and a two-axis emitting Tx, employing a dipole model for position estimation. Cuong *et al.* [26] conducted a comparative study of various electromagnetic models and applied nonlinear optimization for pose estimation. However, these methods mainly rely on nonlinear optimization algorithms to solve system models. This typically results in increased error because such optimization algorithms often face challenges like local minima, sensitivity to initial values, and suboptimal computation time. In particular, in real-time tracking problems where both position and orientation of the capsule change continuously, tracking the actual trajectory can lead the algorithm to converge only to a local optimum.

Some studies using neural networks have attempted to address these limitations. Sasaki *et al.* [27] proposed a machine learning-based model for position localization and compared it with traditional optimization algorithms. Qiang *et al.* [28] applied a backpropagation neural network to estimate the position of the WCE. Qi *et al.* [29] used the levenberg–marquardt optimization method combined with particle swarm optimization, which does not require specific initialization values. Zibo *et al.* [30] proposed a deep learning model based on kernel attention networks (KAN) to replace traditional methods. How-

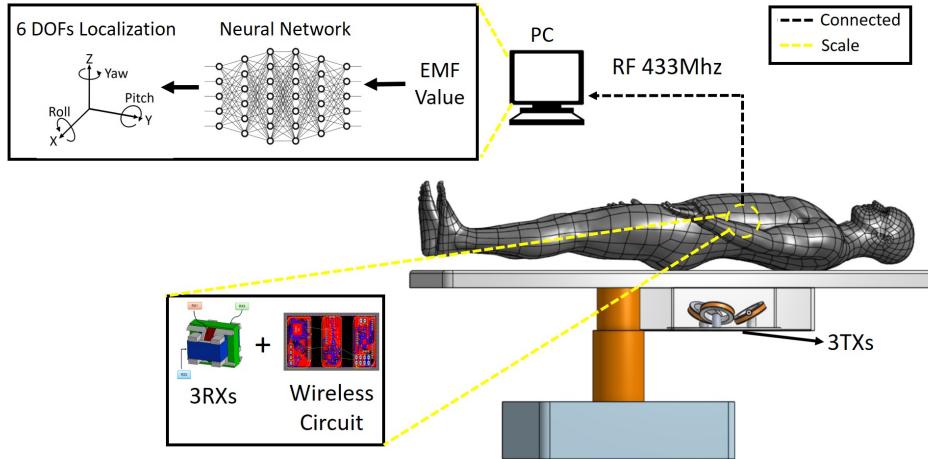


Fig. 1. Overview flowchart of the proposed capsule endoscope localization system based on electromagnetic induction.

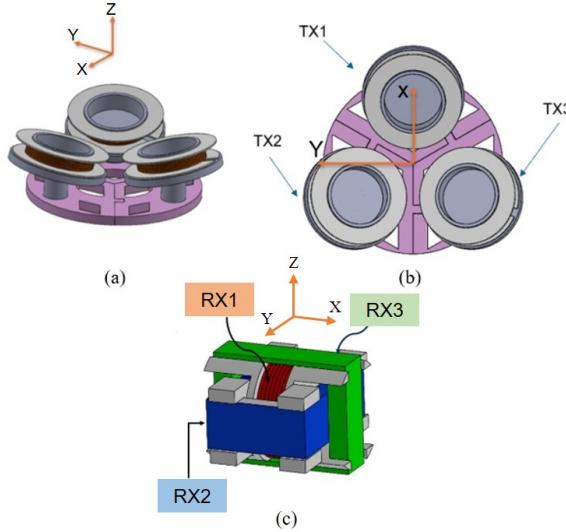


Fig. 2. (a) Configuration of TX with coordinate systems. (b) Top-down view of the TX. (c) Configuration of RX.

ever, most of these methods only evaluate position errors when the rotation angle of the capsule endoscope is fixed or do not assess the orientation error when the capsule is continuously moving in real-world scenarios. This is because continuous variation in the rotation angle during position tracking causes large fluctuations in position error. This highlights the need for a reliable inverse mapping algorithm capable of estimating the 6-DOF pose of the capsule endoscope from EMF signals with high accuracy and robustness to noise, particularly under dynamic and real-time conditions.

In this paper, we propose a 6-DOF capsule endoscope pose estimation algorithm based on a neural network combined with residual blocks. Our proposed algorithm is capable of accurately approximating the system's nonlinear equations, making it a potential substitute for traditional optimization-based methods. The overall system overview is illustrated in Fig 1. The proposed system consists of 3 orthogonally arranged TXs and 3 orthogonally arranged RXs embedded in

TABLE I
SPECIFICATIONS OF TX COILS

TX	Position (cm)	Orientation (α, β, γ)	Turns	Input (V)
TX1	(8.50, 0.00, 7.25)	0°, -15°, 0°	300	20
TX2	(-4.25, 7.36, 7.25)	15°, 0°, 30°	300	20
TX3	(-4.25, -7.36, 7.25)	15°, 0°, 120°	300	20

the WCE, as shown in Fig. 2(c), where the TXs are driven by AC current signals with different frequencies. A localization system is fabricated to validate the feasibility of the algorithm through experiments in which both the angle and the position vary as the capsule moves. A compact wireless circuit is connected to the RX coils to collect EMF signals from the RXs. This circuit then transmits the EMF data to a central computer, where localization algorithms are used to calculate the 6-DOF pose of the capsule endoscope. A 6-DOF robotic arm is also connected to the computer, and the RX coil along with the wireless circuit is mounted on the robotic arm to carry out motion experiments as well as serve as a reference for the error evaluation process. Experimental results show that the proposed method can accurately track the 6-DOF pose of the WCE during real-time movement, with significantly reduced computation time.

II. PROPOSED DEEP LEARNING BASE MAGNETIC FIELD EQUATION

A. Modeling methods of magnetic induction-based localization

The hardware of the capsule positioning system based on the principle of electromagnetic induction consists of transmitting coils and receiving coils. The transmitting coils (TX) are supplied with AC power to create a changing magnetic field in a certain space. When the receiving coils are in this space, they will be induced electromotive forces under the influence of electromagnetic induction. In the proposed hardware, we build 3 transmitting coils and 3 receiving coils with the parameters shown in Table I. The position of WCE is denoted $\mathbf{P} = [X, Y, Z]^T$, and the orientation angles (α, β, γ)

about the X, Y, and Z axes.

Modeling the System: Based on Faraday's law of induction, the electromotive force at the pickup coil is calculated as follows:

$$\varepsilon_k = -\mathbf{n}_k \cdot \frac{d\Phi_k^{(l)}}{dt} = -\mathbf{n}_k \cdot \left(2\pi f_l \Phi_k^{(l)} \cos(2\pi f_l t) \right) \quad (1)$$

where $k (= 1, 2, 3)$ and $l (= 1, 2, 3)$ denote the k -th RX and l -th TX, respectively; n_k is the number of turns of the k -th RX, $\Phi_k^{(l)}$ is the magnetic flux from the l -th TX passing through the k -th RX, and f_l is the frequency of the input current for the l -th TX. The induced voltage has the same frequency as the input current and is 90 degrees out of phase. The amplitude of this voltage depends on the magnetic flux passing through the cross section of the RX receiver coil. This magnetic flux is denoted as $\Phi_k^{(l)}$, which is determined by projecting the magnetic field vector onto the normal direction of the receiver coil, according to the following expression:

$$\Phi_k^{(l)} = A_k \cdot \frac{\mathbf{B}_l(\mathbf{P}) \cdot \mathbf{R}\mathbf{x}_k}{\|\mathbf{R}\mathbf{x}_k\|} \quad (2)$$

where A_k is the cross sectional area of the k -th RX coil, $\mathbf{B}_l(\mathbf{P})$ is the magnetic field vector received at the position of the capsule \mathbf{P} generated by the l -th TX coil, and $\mathbf{R}\mathbf{x}_k$ is the orientation vector of the k -th RX coil. The receiver coils are placed inside the capsule and their orientation can be calculated based on the rotation vector matrix as follows:

$$\mathbf{R}\mathbf{x}_k = \mathbf{R}_{CE} \hat{\mathbf{R}}\mathbf{x}_k \quad (3)$$

$$\mathbf{R}_{CE} = \text{Rot}_z(\gamma) \text{Rot}_y(\beta) \text{Rot}_x(\alpha) \quad (4)$$

where \mathbf{R}_{CE} denotes the rotation matrix of the capsule. The functions $\text{Rot}_z(\cdot)$, $\text{Rot}_y(\cdot)$, and $\text{Rot}_x(\cdot)$ denote rotation matrix operators about the z -, y -, and x -axes, respectively. The vector $\hat{\mathbf{R}}\mathbf{x}_k$ represents the initial direction of the k -th receiver (RX) coil, where $\hat{\mathbf{R}}\mathbf{x}_1 = [1, 0, 0]^T$, $\hat{\mathbf{R}}\mathbf{x}_2 = [0, 1, 0]^T$, and $\hat{\mathbf{R}}\mathbf{x}_3 = [0, 0, 1]^T$.

In this paper, in order to shorten the computation time during training, we use the point-dipole model to estimate the magnetic flux density, instead of applying the Biot–Savart law which requires complex integration [11].

$$\mathbf{B}_l(\mathbf{P}) = \frac{\mu_0 M_l}{4\pi} \left(3 \frac{(\mathbf{D} \cdot \boldsymbol{\Pi}) \boldsymbol{\Pi}}{\|\boldsymbol{\Pi}\|^5} - \frac{\mathbf{D}}{\|\boldsymbol{\Pi}\|^3} \right) \quad (5)$$

$$\mathbf{D} = \mathbf{R}_L^W \cdot [0 \ 0 \ 1]^T \quad (6)$$

where \mathbf{P}_{L_f} is the position of the l -th TX, $\boldsymbol{\Pi} = \mathbf{P} - \mathbf{P}_{L_f}$ is the vector connecting the TX coil and the RX, \mathbf{D} is the moment direction. The peak electromotive force signals are expressed by follow:

$$\left(\varepsilon_k^{(l)} \right)_p = n_k \cdot 2\pi f_l \cdot A_k \cdot \left(\mathbf{B}_l(\mathbf{P}) \cdot (\mathbf{R}_{CE} \hat{\mathbf{R}}\mathbf{x}_k) \right) \quad (7)$$

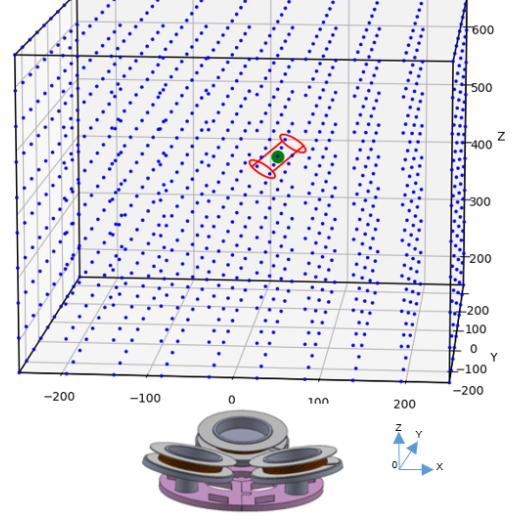


Fig. 3. The localization space is uniformly discretized over the entire rectangular volume, and each point is further assigned a set of uniformly spaced orientation angles

B. Deep learning model

Based on the modeling system in part II-A, with a system configuration consisting of 3TXs and 3RXs, each transmitter-receiver pair yields one equation. Therefore, for each specific position of the WCE, 9 electromotive force (EMF) equations are obtained. We define two problems as follows:

- **Forward Problem:** Starting from the 6-DOF pose of the WCE, and using the model in part II-A, we can easily compute the corresponding 9 EMF signals, since this is a one-to-one mapping.
- **Inverse Problem:** This is the actual problem in practice, where 9 EMF signals are received from the 3 RX coils, and the goal is to compute the 6-DOF pose of the WCE. This is a nonlinear problem, and thus requires powerful algorithms to effectively solve this inverse mapping.

To solve the inverse problem, previous studies employed traditional nonlinear optimization algorithms [21], [24], [26]; however, these methods heavily depend on the initialization parameters and are prone to falling into local minima when the initial value is far from the global minimum. This results in a discrepancy between the predicted and reference values, and significantly increases computation time. Several recent methods have been proposed to address this issue [27]–[30]; however, these studies typically only evaluate the system when the WCE moves under fixed orientation conditions along a predefined trajectory. This does not accurately reflect real-world scenarios, as the system model equations depend on both orientation and position. If the orientation changes continuously, the resulting positional error also varies accordingly. To overcome this limitation, we propose a neural network model combined with residual blocks as the algorithm to address it. The model is enhanced with a residual block, enabling it to learn orientation values more effectively than the feedforward model proposed in previous research [28]. Our goal is to construct a model capable of accurately approximating the

TABLE II
GRID DIVISION AND RANGE OF COORDINATES AND DIRECTIONS

Parameter	Position (X, Y)	Position (Z)	Rotation $\cos(\alpha), \cos(\beta), \cos(\gamma)$
Grid point	40 points	40 points	10 points
Range	[-250, 250] (mm)	[150, 650] (mm)	[-1, 1]

inverse function. In part 1, we first focus on building the dataset that the model will use for training. Then, in part 2, we analyze the structure of the proposed model.

1) *Dataset*: To construct the dataset, we utilize the forward problem, which is a one-to-one mapping problem and, therefore, easy to solve. The input of the problem is the position and orientation of the WCE, and the output consists of 9 EMF values corresponding to the 3 RX coils.

First, we constrain the localization region. Since RXs are symmetrically structured along the three rotation axes, we only consider rotations within the range $\theta \in [0, \pi]$. Since the human abdominal region typically ranges from 150–300 mm in each direction [24], we limit the dataset to a 500 mm side length cube to ensure full coverage of the region of interest while minimizing computational costs.

Next, we divided the space into a uniform grid within a rectangular cuboid region, as illustrated in Fig. 3. At each point on the grid, the orientation angles are also uniformly sampled. Each combination of position and orientation on the grid is then used as input for the forward problem to compute the corresponding EMF values. By performing this computation across the entire grid, we generate a dataset consisting of 6-DOF position and orientation entries, along with the corresponding 9 EMF values.

Finally, instead of using the original variables (α, β, γ) , we propose using new variables $(\cos(\alpha), \cos(\beta), \cos(\gamma))$. In this case, the forward equation takes the form $(\varepsilon_k^l)_p = f(x, y, z, \cos(\alpha), \cos(\beta), \cos(\gamma))$, which helps reduce the complexity of the inverse problem. Depending on a given value of $\cos(\theta)$, only one unique value of θ will be obtained. This is demonstrated by cosine function being a one-to-one mapping over the interval $[0, \pi]$. Thus, each value of $\cos(\theta)$ within this range corresponds to a unique value of θ . We place the origin of the coordinate system at the center of the three transmitting coils, with the axes OX , OY , and OZ defined as illustrated in Fig. 2. The entire localization space is a cube with a side length of 500 mm, and the cosine values of the rotation angles are within the range $[-1, 1]$, corresponding to $\theta \in [0, \pi]$. Table II describes the grid division parameters within the localization space.

2) *Proposed Model*: A neural network is a mathematical model constructed based on the structure of biological neurons [28], in which each neuron receives input and performs a nonlinear transformation before passing the result to subsequent layers. These networks are capable of approximating arbitrary nonlinear functions through linear transformations and nonlinear activation functions, which are trained using backpropagation and gradient descent algorithms [31]. Given an input vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, the output of a single

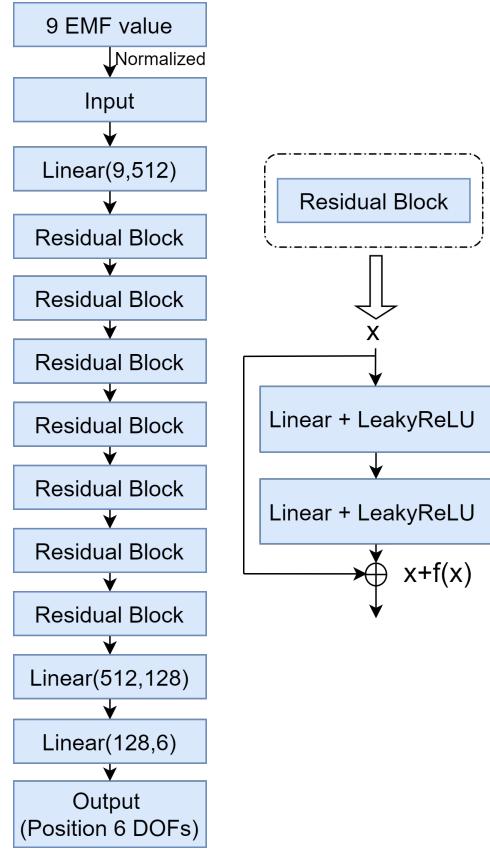


Fig. 4. The proposed neural network model with residual blocks.

neuron is computed as:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right), \quad (8)$$

where w_i is the weight, b is the bias term, and $f(\cdot)$ is the activation function.

However, when training deep networks, performance degradation often occurs as the number of layers increases, leading to a rise in training error. This is not due to overfitting, but rather results from difficulties encountered during gradient descent optimization. To address this issue, a neural network model incorporating residual blocks was introduced [32]. In this model, instead of learning the direct mapping $H(x)$, the network is designed to learn the residual function $F(x) := H(x) - x$, with the original function expressed as:

$$H(x) = F(x) + x. \quad (9)$$

This structure is built with multiple residual blocks that allow for shortcut connections, bypassing several layers and directly adding their input to the output.

Our model is designed to approximate the nonlinear system of equations developed in the previous sections. The proposed architecture consists of multiple residual blocks, as illustrated in Fig. 4. Initially, the input EMF signals are normalized to follow a standard Gaussian distribution, defined as:

$$\hat{x}_i = \frac{x_i - \mu_i}{\sigma_i}, \quad \hat{x}_i \sim \mathcal{N}(0, 1) \quad (10)$$

TABLE III
PARAMETERS OF LAYERS IN EACH RESIDUAL BLOCK

Layer	Parameter
Linear	512×512
LeakyReLU	Slope = 0.01

where x_i is the i -th raw EMF signal, μ_i and σ_i are the mean and standard deviation of that signal across the dataset, and \hat{x}_i is the normalized value which follows the standard normal distribution $\mathcal{N}(0, 1)$.

These normalized features are then passed through a linear layer followed by a series of residual blocks. Each residual block comprises two linear layers, with two nonlinear LeakyReLU activation functions inserted after each linear layer. The LeakyReLU function allows gradients to be maintained even for negative input values. Moreover, since ReLU is a piecewise linear function, its computation time is faster compared to other activation functions such as Tanh or Sigmoid.

The LeakyReLU activation function used in the residual blocks is defined as:

$$\text{LeakyReLU}(z) = \begin{cases} z, & \text{if } z \geq 0 \\ \alpha z, & \text{if } z < 0 \end{cases} \quad (11)$$

where α is a small constant (typically $\alpha = 0.01$) that controls the slope of the negative part of the function. This activation function helps preserve small gradient flows even for negative inputs, improving training stability and convergence. Each residual block acts as a shortcut connection that allows the gradient to flow directly through the layers during backpropagation, helping to mitigate the vanishing gradient problem when training deep networks. The parameters of the layers within each residual block are detailed in Table III. For the loss function, we use the mean squared error (MSE) between the predicted and ground truth localization positions. The MSE is defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{p}_i - \hat{\mathbf{p}}_i\|^2 \quad (12)$$

where \mathbf{p}_i denotes the ground truth 6-DOF pose vector of the capsule, $\hat{\mathbf{p}}_i$ is the corresponding predicted position, and N is the total number of samples.

III. EXPERIMENTAL RESULT

A. Evaluation of the proposed model

We conducted two comparisons on convergence speed over 600 epochs. First, partitioning the dataset based on angle and based on the cosine of the angle. Second, the comparison involves the FCN model and the Full Connected Network (FCN) model combined with residual blocks. In this context, the FCN model has the same configuration of linear layers and LeakyReLU activations as our proposed model; however, it does not include the residual block shortcut connections. Additionally, we also highlighted the performance at every 50-epoch interval. We use the dataset described in Section II.B.2 to train the model, and simultaneously use a test dataset

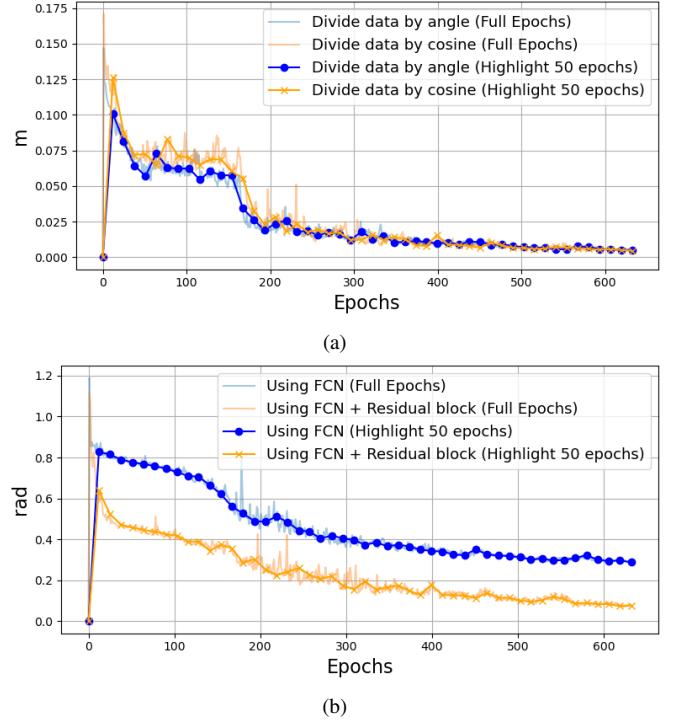


Fig. 5. Comparison of convergence when partitioning angular data by cosine. (a) Convergence rate of position. (b) Convergence rate of orientation.

consisting of 1000 random points with $x, y \in [-250, 250]$ (mm), $z \in [150, 650]$ (mm), and $\cos(\alpha), \cos(\beta), \cos(\gamma) \in [0, 1]$. This dataset is passed through the forward problem to calculate the EMF values, which are then fed into the model to infer the predicted positions. The predicted results are then compared with the reference values corresponding to the 1000 random points above. To compare the convergence speed of the models, we evaluate and compare the convergence rate of position and orientation errors on the test dataset. The formulas for calculating position and orientation errors were defined:

$$E_{\text{pos}} = \sqrt{\frac{\|\mathbf{P}_T - \mathbf{P}_R\|^2}{3}} \quad (13)$$

$$E_{\text{dir}} = \sqrt{\frac{(\alpha_T - \alpha_R)^2 + (\beta_T - \beta_R)^2 + (\gamma_T - \gamma_R)^2}{3}} \quad (14)$$

where E_{pos} and E_{dir} represent the root mean square values of the position error and orientation error, respectively. $(\mathbf{P}_T, \alpha_T, \beta_T, \gamma_T)$ are the obtained position and orientation results, and $(\mathbf{P}_R, \alpha_R, \beta_R, \gamma_R)$ are the reference values. The entire training and inference process was conducted on a central computer equipped with an Intel(R) Core(TM) i9-10980XE 3.00GHz processor and an RTX 3090 GPU.

1) *Comparison between splitting the dataset based on the angle and based on the cosine of the angle:* The results are presented in Fig. 5. The partitioning of the dataset based on cosine similarity results in a smoother and more uniform distribution of the input, thereby improving the convergence process for the angular component, which contributes more to the nonlinearity. Meanwhile, the convergence speed of the position, which contributes less nonlinearity to the equation, almost reaches a steady-state value after 600 epochs.

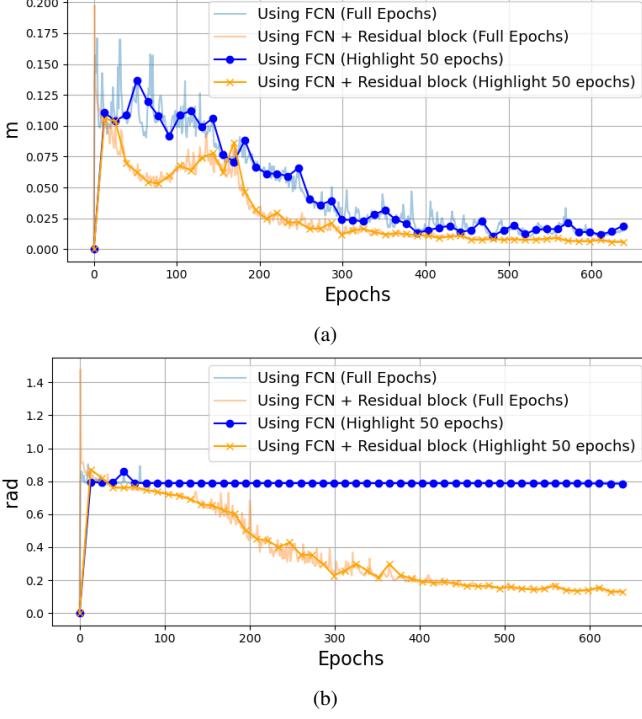


Fig. 6. Comparison of convergence speed between FCN and FCN combined with residual blocks. (a) Convergence speed of position. (b) Convergence speed of orientation.

2) Comparison between FCN and FCN with Residual Blocks: In Fig. 6, the convergence speed between the FCN model and the FCN model combined with residual blocks is analyzed. The residual blocks create shortcut connections between layers, helping to prevent gradient vanishing when the model becomes deeper. The results show that for the angular component, our model demonstrates significantly better learning capability, while the standard FCN model exhibits very slow convergence even after 600 epochs. For the position component, our model also achieves faster convergence compared to the conventional FCN model.

B. Comparison of the system's ability to approximate the nonlinear equation between our proposed approach and nonlinear optimization algorithms.

This section compares the ability to approximate the nonlinear equations of the inverse problem between our proposed method and the conventional nonlinear optimization algorithm. The input consists of 9 EMF values, and the output is the 6-DOF pose of the wireless capsule endoscope. To establish a reference, we first use the forward problem: a dataset of known 6-DOF pose within the localization space is fed into the forward model to compute the corresponding 9 EMF values. These EMF values then serve as the input for the inverse problem. To evaluate the nonlinear approximation capability of the inverse problem, the EMF values generated by the forward model are input into both our proposed method and the traditional nonlinear optimization algorithm. Each method produces an estimated 6-DOF pose. These estimated values are then compared against the original 6-DOF pose

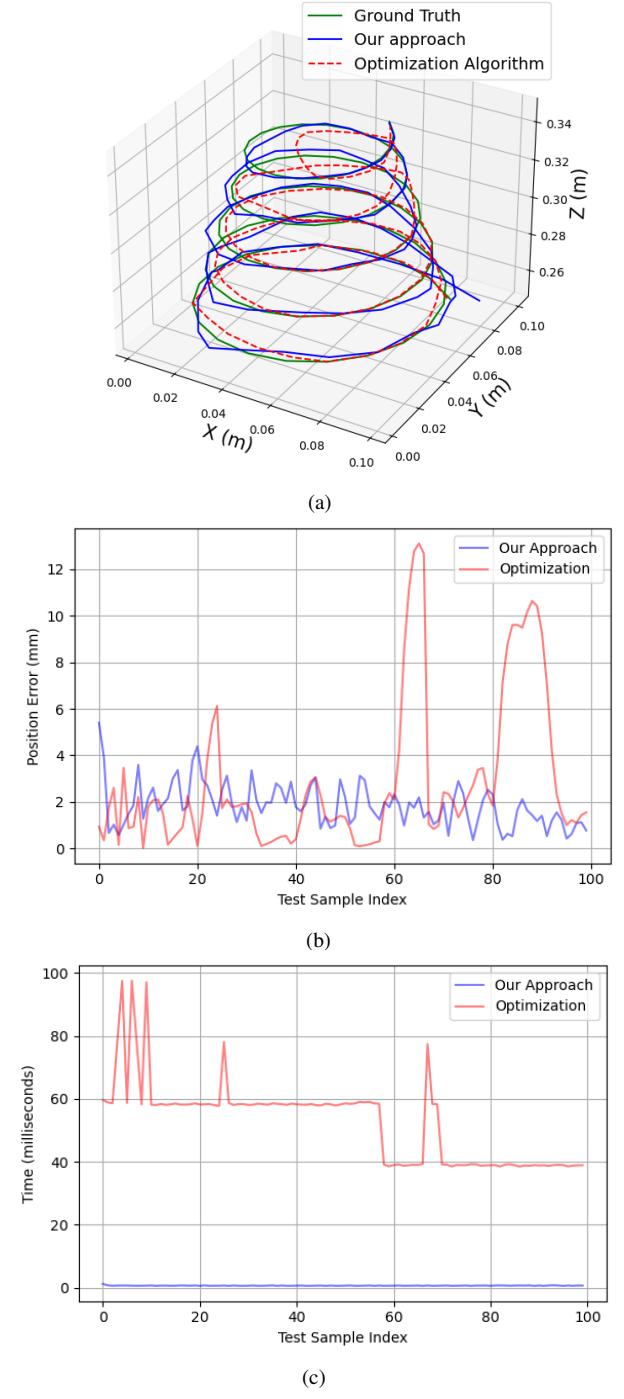


Fig. 7. Comparison of the nonlinear equation approximation capability between our proposed method and a traditional optimization algorithm with a helix trajectory and fixed roll, pitch, and yaw angles. (a) Position error. (b) Computation time.

from the forward problem to compute the error and assess the approximation accuracy of each approach. In this study, we utilize a nonlinear optimization algorithm based on the `least_squares` function from the SciPy library, employing the Trust Region Reflective solver. The objective of the algorithm is to minimize the loss function, which represents the squared error between the experimentally measured EMF data and the EMF values computed from the forward model,

TABLE IV
COMPARISON OF POSITION RMSE AND COMPUTATION TIME USING A HELIX TRAJECTORY WITH FIXED ROTATION

Method	RMSE (mm)	Computation time (ms)
Our method	1.85	0.86
Optimization	2.88	52.75

subject to boundary constraints defined according to each experimental scenario. To evaluate the system's performance, a three-axis system with self-adjusting rotating axes was used to move the RCE along a 3D helical trajectory in various directions. To enable quantitative performance comparison, the position and orientation errors are defined as shown in (13) and (14). The central computer is equipped with an Intel(R) Core(TM) i9-10980XE 3.00GHz processor.

We conducted three evaluation scenarios to compare our approach with a nonlinear optimization algorithm. First, we performed a comparison using a helix trajectory with fixed roll, pitch, and yaw angles. Next, we evaluated the methods using a randomly generated trajectory with both varying positions and orientations. Finally, we compared the performance on a helix trajectory with continuously changing roll, pitch, and yaw angles.

1) *A helix trajectory with fixed roll, pitch, and yaw angles:* We use the datatest set, which is a helix trajectory consisting of 5 loops, with the roll, pitch, and yaw angles fixed at 60 degrees

The results are shown in Figure 7, and Table IV. When the roll, pitch, and yaw angles are kept fixed and the trajectory consists of consecutive neighboring points, the initial point is close to the global optimum. Therefore, the algorithm achieves a low error of 2.88 mm, and the computation time is reduced to only 52.75 ms. For our approach, the error achieved is 1.82 mm, while the computation time is lower than the optimization algorithm, taking only 0.86 ms. These results show that the optimization algorithm can maintain a low position error under the condition that the rotation angles are constant and the trajectory consists of successive neighboring points.

2) *Randomly trajectory:* We used a test dataset of 1000 random points, with both the rotation angle and position randomly changed without following any specific pattern. At the same time, we conducted a comparison with traditional optimization methods. The test space was defined as: $X, Y \in [-250, 250]$ (mm), $Z \in [150, 650]$ (mm), $\cos(\alpha), \cos(\beta), \cos(\gamma) \in [0, 1]$.

The results, as presented in Fig. 8 and Table V, show that when the trajectory consists of randomly distributed points, the localization error increases significantly, and the computation time also rises. This is because, with a random distribution of the trajectory, the initial point of the algorithm cannot be set close to the global optimal solution. As a result, the optimization algorithm falls into local minima, causing both the error and computation time to increase substantially. In a practical system, each step of data acquisition often takes considerable time due to factors such as wireless communication and computation, so the trajectory points at each step are usually spaced far apart. This leads to increased system error,

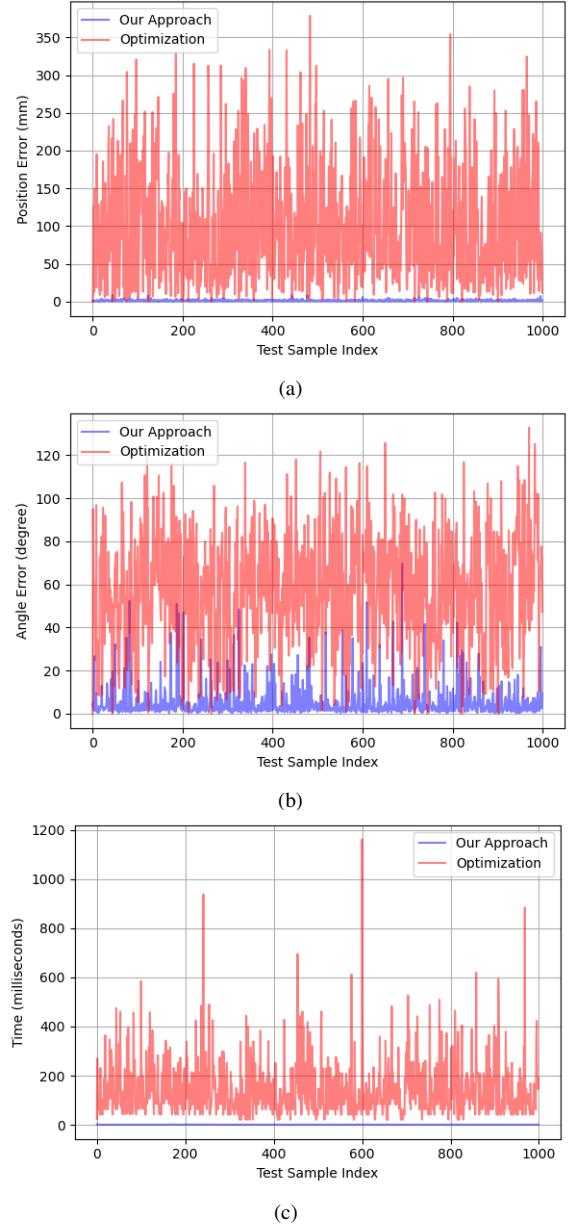


Fig. 8. Comparison of nonlinear equation approximation capability between our proposed method and traditional optimization algorithm with randomly trajectory. (a) Position error. (b) Orientation error. (c) Computation time.

TABLE V
COMPARISON OF POSE ESTIMATION ACCURACY AND COMPUTATION TIME USING A RANDOM TRAJECTORY

Method	RMSE (mm)	RMSE of Angle ($^{\circ}$)	Time (ms)
Our method	1.64	5.15	0.86
Optimization	101.75	57.81	142.23

as the optimization algorithm may experience a significant rise in error under such conditions.

3) *A helix trajectory with randomly varying roll, pitch, and yaw angles:* In this section, we conduct a comparison between our proposed algorithm and the nonlinear optimization method using a trajectory composed of consecutive points along a predefined spiral path with a maximum diameter of 500 mm, as illustrated in Fig. 8(a) and $\cos(\alpha), \cos(\beta), \cos(\gamma) \in [0, 1]$.

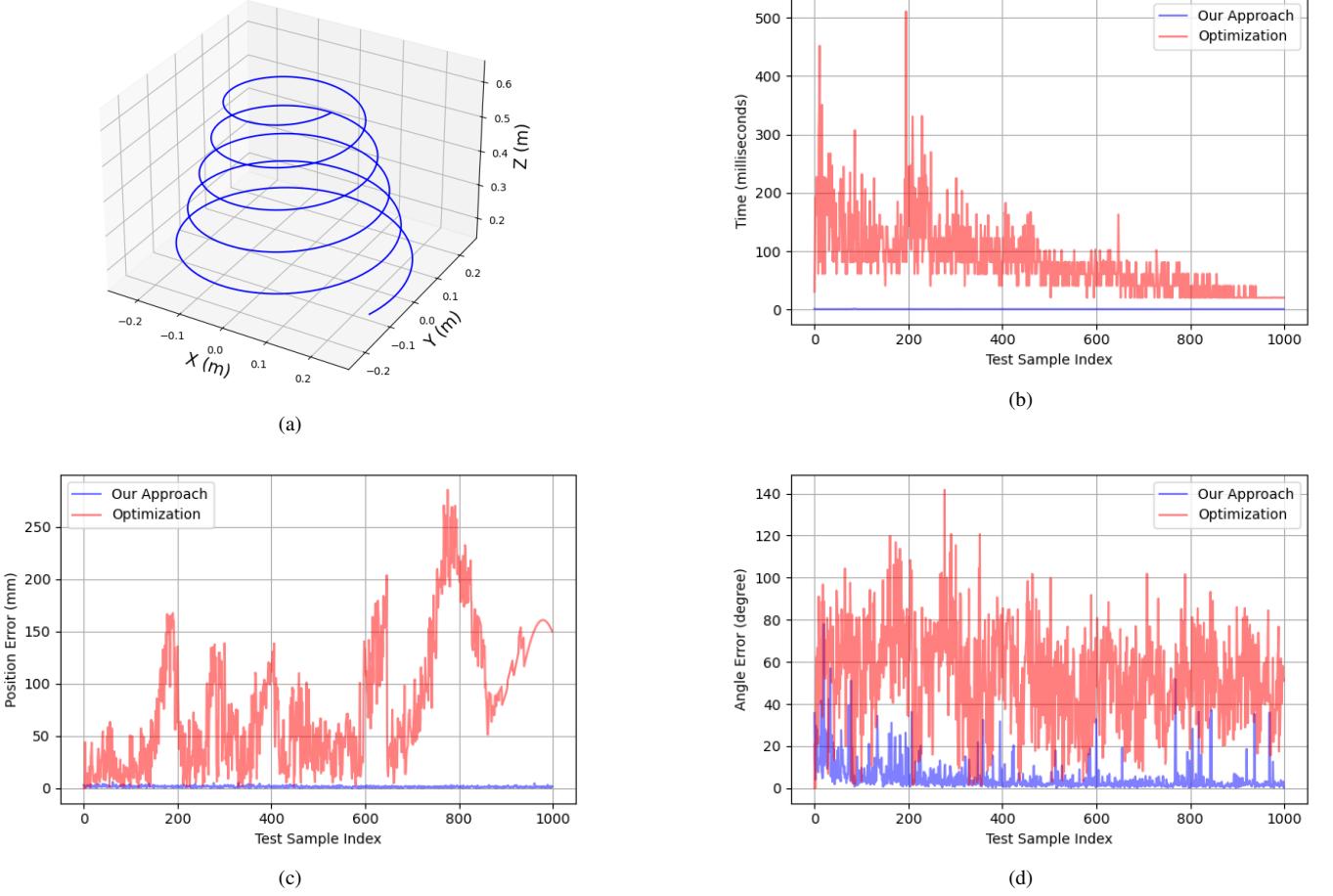


Fig. 9. Comparison of nonlinear equation approximation capability between our proposed method and traditional optimization algorithm using a sequential helix trajectory. (a) Helix trajectory. (b) Computation time. (c) Position error. (d) Orientation error.

TABLE VI
COMPARISON OF POSE ESTIMATION ACCURACY AND COMPUTATION TIME USING A HELIX TRAJECTORY WITH RANDOM ROTATION

Method	RMS Error Position (mm)	RMS Error Angle (°)	Time (ms)
Proposed Method	1.54	4.81	0.76
Optimization Algorithm	60.91	52.71	82.28

During the movement, the rotational angles are also continuously varied to evaluate the error between the two models.

The results presented in Fig. 9(a), (b), and (c), and Table VI show that even when both the position and the angle continuously vary and although consecutive points are used to facilitate global optimum convergence for the optimization algorithm, which still yields significantly larger errors. This is primarily because the angle is a highly nonlinear component. In previous studies, the angular values were often kept fixed while the WCE moved to reduce the overall error. However, in our proposed method, despite simultaneous variations in both position and angle, the model achieves low error in both components while also significantly reducing computation time. This robustness is attributed to the fact that our model does not rely on an initial guess, thereby maintaining low

errors across varying conditions.

C. Experiment with the Real System.

1) *Experiment Setup:* The entire experimental setup process is illustrated in Fig. 10. The experiment setup consists of three TX coils connected to three function generators (PeakTech 3 MHz DDS 4055MV) to supply power. The signal generators are configured at frequencies of 4000 Hz, 4500 Hz, and 5000 Hz, respectively, with a peak-to-peak voltage amplitude of 5 V. The RX coils have cross-sectional areas of 3π mm², 10×7 mm², and 10×10 mm², with 250, 200, and 200 turns, respectively.

To address the limitations of previous studies [24]–[26] that used wired connections to transmit the EMF signals from the RX coils to the central computer for processing, we designed a prototype circuit board with dimensions of 22 mm × 12 mm, which is stacked and embedded inside the WCE. Fig. 11 shows the image of the hardware diagram and prototype circuit board. The board includes a Sub-1 GHz transceiver IC from Texas Instruments, an STM32F411CEU6 microcontroller with an ARM Cortex-M4 core, an RT0913 buck converter regulating the voltage to 3.3 V, and a 16-bit 250 kSps ADC (AD7689) from Analog Devices. In addition,

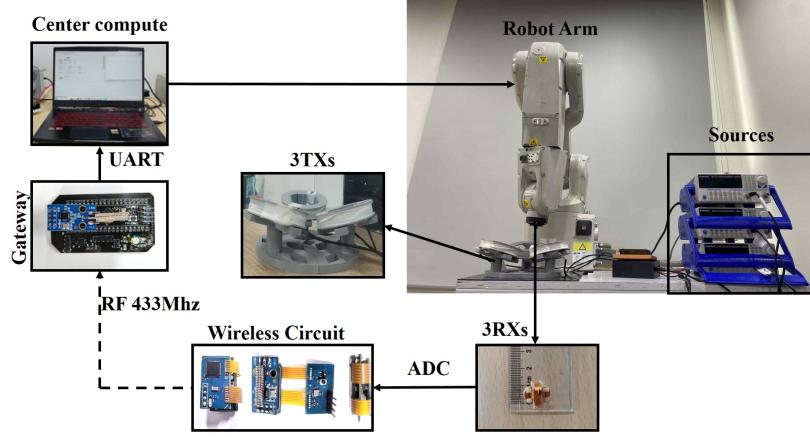


Fig. 10. The experimental setup for evaluating the positional and angular errors of the endoscopic capsule system.

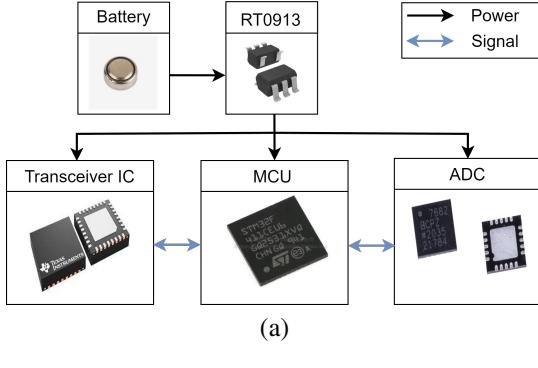


Fig. 11. (a) Hardware diagram; (b) Assembled flexible modular prototype.

we designed an external gateway module that uses a Sub-1 GHz transceiver IC from Texas Instruments, which is capable of communicating with the internal prototype board inside the WCE via an RF433 MHz link. This gateway is connected to the central computer through a UART interface to transmit the data for processing. Simultaneously, the central computer is also connected to an ABB robotic manipulator, which is used to provide reference motion for performance evaluation. The central computer is equipped with an Intel(R) Core(TM) i9-10980XE 3.00 GHz processor and an RTX 3090 GPU.

2) *Calibration:* In practice, when using the mathematical model equations, the system may encounter various problems that cause errors, such as discrepancies in the positions and angles of the transmitter coils between the design and the actual assembly due to human factors, and deviations in the number of turns of the RX and TX coils. These parameters are considered fixed in both the forward and inverse problems, so any deviations will affect the actual positioning error of the

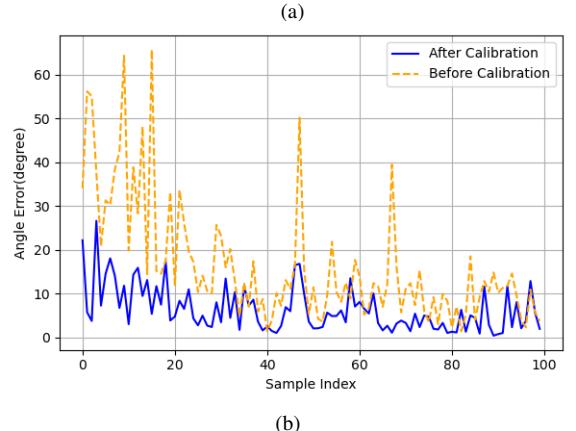
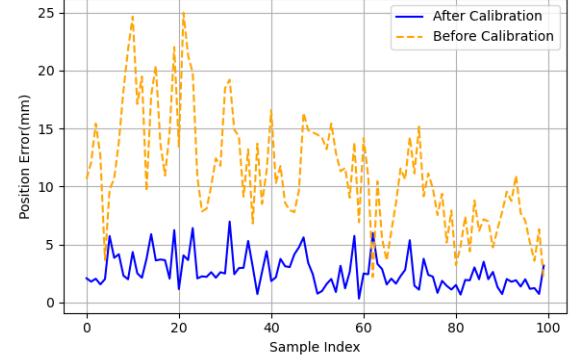


Fig. 12. (a) Positional error before and after calibration. (b) Angular error before and after calibration.

system.

To solve this problem, we propose a two-step calibration method. In the first step, we collect 150 random real-world points where $X, Y \in [-250, 250]$ mm, $Z \in [150, 650]$ mm, $\cos(\alpha), \cos(\beta), \cos(\gamma) \in [-1, 1]$. After that, we use an optimization tool to solve the least squares optimization problem using the mathematical model in Section II, where the inputs are the positions, orientations, and corresponding EMF values of these 150 points, and the outputs of the system are the values of the angles and directions of the three transmitter coils, as well as the number of turns of the

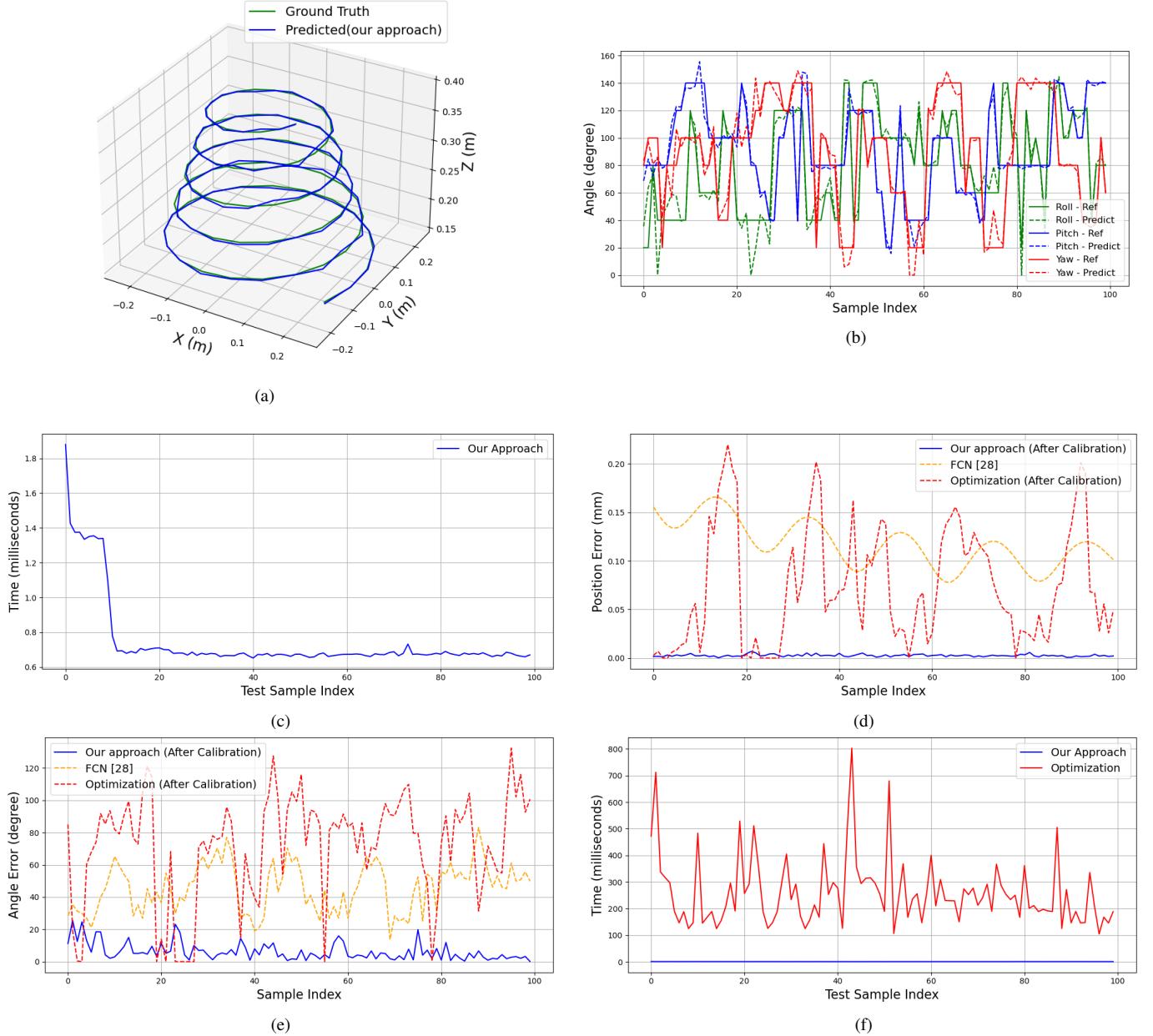


Fig. 13. Experimental tracking results of the capsule's position. (a) Position tracking. (b) Orientation tracking. (c) Computation time. Comparison between our proposed method and the optimization algorithm in a real-world system. (d) Position error. (e) Orientation error. (f) Computation time.

transmitter coils that match the actual system. In the second step, the fixed parameters such as the values of the angles, directions of the TX coils, and the number of turns of the transmitter coils are updated for both the forward and inverse problems to retrain the model as in Section II.B.

The test dataset used for evaluating the RMS errors of position and orientation according to formulas (13) and (14), before and after calibration, follows a helix path made up of five loops, with the largest loop having a radius of 150 mm and a height of 200 mm. The three rotation angles, roll, pitch, and yaw, are continuously varied from 20° to 160° in steps of 20° , consisting of a total of 100 points. The results of the calibration process are shown in Fig. 12. After calibration, both position and orientation errors were significantly reduced. Specifically,

TABLE VII
COMPARISON OF RMSE IN POSITION AND DIRECTION

Method	RMSE	
	Position (mm)	Direction ($^\circ$)
Optimization Algorithm	48.85	68.89
FCN [28]	101.62	40.68
Our Approach	1.90	3.55

the RMSE of position decreased from 8.42 mm to 1.90 mm, while the RMSE of orientation was reduced from 12.02° to 3.55° .

3) *Localization Results and Comparison Between Our Approach, Optimization Algorithm, and FCN [28]:* To demonstrate the performance of the proposed neural network model

TABLE VIII
COMPARISON OF THE PROPOSED APPROACH WITH EXISTING POSITIONING METHODS FOR CAPSULE ENDOSCOPE

Parameters	Our Approach	Qiang <i>et al.</i> [28]	Qi <i>et al.</i> [29]	Zibo <i>et al.</i> [30]	Cuong <i>et al.</i> [26]	Liu <i>et al.</i> [24]
Localization method	FCN + Residual Block	FCN	L-M optimization + PPO	KAN	Nonlinear Optimization	Nonlinear Optimization
Degrees of freedom	6	3	5	5	6	6
Dynamic error test	Yes	No	No (Fix angle)	No (Fix angle)	No (Fix angle)	No (Fix pitch and yaw)
Evaluation Area	500×500×500 mm, $2\pi \times 2\pi \times 2\pi$	100×100×100 mm	150×150×150 mm, 15°–75°	300×180×180 mm	100×100×100 mm	Sphere (100 mm diameter)
Trajectory	Helix and random path	Circle and lines on 2D plane	Circle and lines on 2D plane	N/A	Square trajectory	Helix path
Position error	1.90 mm	0.01 mm	1.32 mm	0.10 mm	2.10 mm	1.68 mm
Direction error	3.55°	N/A	1.33°	N/A	3.80°	1.74°
Computation time	0.82 ms	N/A	13 ms	N/A	14 ms	200ms

using residual block, we compared with the model proposed in [28], which has three hidden layers, with the number of neurons in the first, second, and third hidden layers being 8, 15, and 9, respectively. We collected 300 random points in the region $X, Y \in [-250, 250]$ mm, $Z \in [150, 450]$ mm, $\cos(\alpha), \cos(\beta), \cos(\gamma) \in [-1, 1]$ to create the datatrain set for FCN model. The test datatest used for evaluation follows a helix path made up of 5 loops, as shown in Fig. 11(a), with the largest loop having a radius of 15 cm and a height of 20 cm. The three rotation angles roll, pitch, and yaw continuously varied over the set continuously varied from 20° to 160° in steps of 20°, consisting of 100 points in total. We then employed this datatest to compare the performance of our proposed method with nonlinear optimization algorithm and the approach in [28]. The position and orientation errors are defined as shown in (13) and (14).

The results are illustrated in Fig. 13 and Table VII. Specifically, Fig. 13(a) and Fig. 13(b) show the tracking performance of our model in terms of both position and orientation, while Fig. 13(c) presents the computation time for each prediction step. Fig. 13(e) and Fig. 13(f) present a direct comparison between our method, FCN model and the nonlinear optimization approach, while Fig. 13(e) compares the computation time among the three methods. It can be observed that our proposed method is capable of maintaining a position error of 1.9 mm and an angular error of 3.5° under various conditions, even when both the orientation and position change continuously. Furthermore, the computation time is significantly reduced to just 0.82 ms. When comparing with the nonlinear optimization approach, it is evident that, in cases where both position and angle vary simultaneously, the optimization algorithm is more prone to being trapped in local minima, leading to a substantial increase in tracking error. Finally, the proposed method in [28] uses a dataset of 300 real-world points for evaluation, which does not ensure that the training data points cover the entire evaluation space. In addition, in study [28], the evaluation was conducted only for trajectories in 3D space with fixed rotation angles and simple trajectories. When the rotation angles change continuously, the position error of the FCN model increases significantly.

4) *Comparison with Existing Methods:* Table VIII provides a comparison between our proposed method and existing approaches. As shown, most previous studies maintained a fixed rotational component during localization. This component is a major contributor to the nonlinearity of the inverse system model. Our study addresses this issue by introducing experiments where both orientation and position vary continuously

throughout the trajectory, which better reflects the realistic motion of a capsule endoscope. Throughout the tracking process, our method achieves a low error of 1.9 mm in position and 3.55° in orientation. Additionally, the computation time is significantly reduced to just 0.8 ms, outperforming previous methods in both accuracy and speed.

IV. CONCLUSION

In this study, we address a critical real-world challenge: maintaining low tracking errors even when both the position and orientation of the capsule vary simultaneously. We also compare our proposed method with traditional nonlinear optimization algorithms in terms of their ability to approximate the nonlinear inverse mapping. The results clearly demonstrate that our method significantly outperforms the optimization approach, which is often susceptible to local minima and highly dependent on the choice of the initial starting point leading to substantial increases in error. Notably, the computation time of our method is drastically reduced to only 0.82 ms per inference, which is a remarkable improvement compared to previous approaches. This highlights the potential of using function approximation to replace traditional optimization techniques in real-time nonlinear systems. Furthermore, our proposed model requires only 15 MB of memory, making it feasible for deployment on embedded processors. This paves a way for developing a fully integrated standalone system without the need for an external computer, thus eliminating the latency incurred from transmitting data back and forth for computation during each tracking cycle. In the future, we aim to further compress and optimize the model for lightweight deployment and explore the use of other advanced AI architectures to improve both performance and inference speed. At the same time, we will reduce the size of the prototype circuit embedded inside the capsule robot.

REFERENCES

- [1] M. Sendoh, K. Ishiyama, and K.-I. Arai, “Fabrication of magnetic actuator for use in a capsule endoscope,” *IEEE Transactions on Magnetics*, vol. 39, no. 5, pp. 3232–3234, Sept. 2003, doi: 10.1109/TMAG.2003.816731.
- [2] J. L. Toennies, G. Tortora, M. Simi, P. Valdastri, and R. J. Webster, “Swallowable medical devices for diagnosis and surgery: The state of the art,” *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 224, no. 7, pp. 1397–1414, Jan. 2010.
- [3] T. D. Than, G. Alici, H. Zhou, and W. Li, “A review of localization systems for robotic endoscopic capsules,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2387–2399, Sept. 2012.
- [4] W. Marlicz *et al.*, “Frontiers of robotic gastroscopy: A comprehensive review of robotic gastrosopes and technologies,” *Cancers (Basel)*, vol. 12, no. 10, p. 2775, 2020.

- [5] M. Sitti *et al.*, “Biomedical applications of untethered mobile milli/microrobots,” *Proceedings of the IEEE*, vol. 103, no. 2, pp. 205–224, Feb. 2015, doi: 10.1109/JPROC.2014.2385105.
- [6] T. Nakamura and A. Terano, “Capsule endoscopy: past, present, and future,” *Journal of Gastroenterology*, vol. 43, pp. 93–99, 2008. [Online]. Available: <https://doi.org/10.1007/s00535-007-2153-6>
- [7] M. C. Hoang *et al.*, “A robotic biopsy endoscope with magnetic 5-DOF locomotion and a retractable biopsy punch,” *Micromachines*, vol. 11, no. 1, p. 98, 2020.
- [8] F. Munoz, G. Alici, and W. Li, “A review of drug delivery systems for capsule endoscopy,” *Advanced Drug Delivery Reviews*, vol. 71, pp. 77–85, 201410.1016/j.addr.2013.12.007.
- [9] K. T. Nguyen and C. S. Kim, “Medical microrobot — A drug delivery capsule endoscope with active locomotion and drug release mechanism: Proof of concept,” *Int. J. Control Autom. Syst.*, vol. 18, no. 1, pp. 65–75, 2020.
- [10] S. P. Woods and T. G. Constantine, “Wireless capsule endoscope for targeted drug delivery: Mechanics and design considerations,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 945–953, Apr. 2013, doi: 10.1109/TBME.2012.2228647.
- [11] G. Bao, L. Mi, Y. Geng, M. Zhou, and K. Pahlavan, “A video-based speed estimation technique for localizing the wireless capsule endoscope inside gastrointestinal tract,” in *Proc. IEEE 36th Int. Conf. EMBC*, 2014, pp. 5615–5618.
- [12] G. Ciuti, A. Menciassi, and P. Dario, “Capsule endoscopy: From current achievements to open challenges,” *IEEE Rev. Biomed. Eng.*, vol. 4, pp. 59–72, 20110.1109/RBME.2011.2171182.
- [13] D. K. Iakovidis *et al.*, “Intelligent visual localization of wireless capsule endoscopes enhanced by color information,” *Comput. Biol. Med.*, vol. 89, pp. 429–440, 2017.
- [14] A. Wang *et al.*, “Wireless capsule endoscopy,” *Gastrointest. Endosc.*, vol. 78, no. 6, pp. 805–815, 2013, doi: 10.1016/j.gie.2013.06.026.
- [15] D. Son, X. Dong, and M. Sitti, “A simultaneous calibration method for magnetic robot localization and actuation systems,” *IEEE Trans. Robot.*, vol. 35, no. 2, pp. 343–352, Apr. 2019.
- [16] D. M. Pham and S. M. Aziz, “A real-time localization system for an endoscopic capsule,” in *Proc. IEEE 9th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP)*, Apr. 2014, pp. 1–6.
- [17] M.-C. Kim, E.-S. Kim, J.-O. Park, E. Choi, and C.-S. Kim, “Robotic localization based on planar cable robot and Hall sensor array applied to magnetic capsule endoscope,” *Sensors*, vol. 20, no. 20, Art. 5728, Oct. 2020.
- [18] C. Di Natali, M. Beccani, and P. Valdastri, “Real-time pose detection for magnetic medical devices,” *IEEE Transactions on Magnetics*, vol. 49, no. 7, pp. 3524–3527, Jul. 2013, doi: 10.1109/TMAG.2013.2240899.
- [19] M. Wang *et al.*, “Multipoint simultaneous tracking of wireless capsule endoscope using magnetic sensor array,” *IEEE Trans. Instrum. Meas.*, vol. 70, Art. no. 7502510, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3075776.
- [20] K. M. Popek, T. Schmid, and J. J. Abbott, “Six-degree-of-freedom localization of an untethered magnetic capsule using a single rotating magnetic dipole,” *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 305–312, Jan. 2017, doi: 10.1109/LRA.2016.2608421.
- [21] H. Liu, J. Su, and S. Song, “Design of a reconfigurable sensor array for localization of magnetic capsule endoscope,” *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–9, 2025, Art. no. 4000609, doi: 10.1109/TIM.2024.3522675.
- [22] M. N. Islam and A. J. Fleming, “A novel and compatible sensing coil for a capsule in wireless capsule endoscopy for real-time localization,” in *Proc. IEEE Sensors*, 2014, pp. 1607–1610.
- [23] M. N. Islam and A. J. Fleming, “An algorithm for transmitter optimization in electromagnetic tracking systems,” *IEEE Trans. Magn.*, vol. 53, no. 8, p. 8002208, 2017.
- [24] S.-L. Liu, J. Kim, A. Hong, J.-O. Park, and C.-S. Kim, “Six-Dimensional Localization of a Robotic Capsule Endoscope Using Magnetoquasistatic Field,” *IEEE Access*, vol. 10, pp. 22865–22874, 2022, doi: 10.1109/ACCESS.2022.3154031.
- [25] S. Song, C. Hu, B. Li, X. Li, and M. Q.-H. Meng, “An electromagnetic localization and orientation method based on rotating magnetic dipole,” *IEEE Transactions on Magnetics*, vol. 49, no. 3, pp. 1274–1277, Mar. 2013, doi: 10.1109/TMAG.2012.2211375.
- [26] M. C. Hoang, J. Kim, J.-O. Park, and C.-S. Kim, “Six-DOF localization using magnetic induction effect for automated locomotion of an active capsule endoscope,” in *Proc. 9th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Seoul, South Korea, 2022, pp. 1–6, doi: 10.1109/BioRob52689.2022.9925464.
- [27] A.-i. Sasaki and E. Ohta, “Magnetic-field-based position sensing using machine learning,” *IEEE Sensors J.*, vol. 20, no. 13, pp. 7292–7302, Jul. 1, 2020, doi: 10.1109/JSEN.2020.2979071.
- [28] 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, pp. 1–9, 2024, Art. no. 4002509, doi: 10.1109/TIM.2023.3341129.
- [29] Q. Zhang, Y. Li, H. Xu, X. Li, and X. Zhang, “Magnetic localization method of capsule endoscope based on hybrid model,” *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, Art. no. 4003710, doi: 10.1109/TIM.2023.3261943.
- [30] Z. Gao and M. Kong, “MP-KAN: An effective magnetic positioning algorithm based on Kolmogorov-Arnold network,” *Measurement*, vol. 243, Art. no. 116248, 2025, doi: 10.1016/j.measurement.2024.116248.
- [31] S. Amari and K. Saito, “Stochastic gradient descent training of multi-layer perceptrons for deep learning,” 1967.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.



Tran Van Luong received the B.S. degree in electrical engineering from the Hanoi University of Sciences and Technology (HUST), Viet Nam, in 2024. He is currently pursuing the M.S. degree with the School of Electrical Engineering, KAIST. His current research interests include deep learning, reinforcement learning, robotics and embedded systems.



Nguyen Xuan Mai received the B.S. degree in electrical engineering from the Hanoi University of Sciences and Technology (HUST), Viet Nam, in 2024. She is currently pursuing the M.S. degree with the School of Electrical Engineering, KAIST. Her current research interests encompass 6G communication systems, antenna beamforming, reinforcement learning algorithms, and real-time embedded systems.



Dong Eui Chang (Member, IEEE) received the B.S. degree in Control and Instrumentation Engineering and the M.S. degree in Electrical Engineering from Seoul National University, South Korea, and the Ph.D. degree from the California Institute of Technology (Caltech), USA. He was an Associate Professor of Applied Mathematics at the University of Waterloo, Canada. Since 2017, he has been with the Korea Advanced Institute of Science and Technology (KAIST), where he is currently a Full Professor in the School of Electrical Engineering. He is the co-author of the book *Deep Neural Networks in a Mathematical Framework* (Springer, 2018). His research interests include control theory, geometric mechanics, and artificial intelligence.



Si Hong Hoang received his B.E and M.E degrees in the School of Electrical Engineering from Hanoi University of Technology, Vietnam, in 1999 and 2001, respectively, and his Ph.D. degree from the University of Ulsan, Korea, in 2010. Now, he is an Associate Professor at the School of Electrical Electronic Engineering, Hanoi University of Science and Technology, Vietnam. His research interests include smart sensors, piezoelectric material, SAW sensors, and AI-based diagnostic application.



Manh Cuong Hoang received the B.S. (2016) in Electrical Engineering and Ph.D. (2021) degrees in Mechanical Engineering from Hanoi University of Science and Technology, and Chonnam National University, respectively. He was a Postdoctoral Researcher at the Korea Institute of Medical Micro-robotics, South Korea. Now he is His current research interests include magnetic medical micro-robot, control, and artificial intelligence.