A Context-Aware GraphSAGE–Transformer Model for LED-Based Indoor Positioning in Narrow Corridors

Original Scientific Paper

Huy Q. Tran*
Nguyen Tat Thanh University,
Robotics and Mechatronics Lab, Faculty of Engineering and Technology,
Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
tqhuy@ntt.edu.vn

Cheolkeun Ha
University of Ulsan,
Robotics and Mechatronics Lab, University of Ulsan, Ulsan City, Republic of Korea
cheolkeun@gmail.com

*Corresponding author

Abstract *– Indoor localization in confined spaces such as hallways presents significant challenges due to limited direct signal access and increased multipath reflections. Improving localization accuracy in narrow and irregularly shaped hallways is quite difficult. This paper proposes a novel context-aware GraphSAGE–Transformer model for indoor visible light localization. A context-based feature weighting mechanism is designed to learn global received signal strength reliability patterns across multiple LEDs. Simultaneously, the model attempts to eliminate unreliable measurements caused by noise and multipath effects. GraphSAGE is then employed to model local spatial correlations among neighboring LEDs, while a Transformer encoder captures long-range and global dependencies through self-attention. Additionally, we utilize the Huber loss to enhance robustness against severe received signal strength outliers in non-line-of-sight environments. Applied in narrow L-shaped and T-shaped hallways with minimal lighting infrastructure, the proposed method helps to achieve a root mean squared error of 0.2 meters, reducing the error by 55% compared to using Transformer alone and by 44% compared to using GraphSAGE alone. Furthermore, experimental results also demonstrate that the model achieves approximately 3 times better accuracy than traditional machine learning models, such as Random Forests and k-Nearest Neighbors. These results confirm the effectiveness of the proposed model in overcoming the inherent limitations of indoor hallway environments.*

Keywords: *Indoor localization, visible light localization, LED, Transformer, GraphSAGE*

Received: February 15, 2026; Received in revised form: May 15, 2026; Accepted: June 3, 2026

1. INTRODUCTION

LED lights have many valuable features that suit visible light positioning (VLP) systems. First, they are energy-efficient and provide good lighting quality, which must not be affected by the positioning function. The lights should not flicker or change color in a way that people can notice. LEDs also support dimming, so the positioning system must work with different brightness levels. There are two main types of white LEDs: red, green, and blue, and the other uses a blue LED with a phosphor coating. In practical applications, both types of LEDs are used. Therefore, the signal used for the positioning process does not rely on color change. The advantage of LEDs is that they can transmit data inside a room without fear of interference. LED lighting systems are now pre-installed in most households. Replacing them can also be done easily and quickly. A special feature is that we can transmit control signals to LEDs through the existing wiring without needing new wiring. This makes LEDs a suitable choice for indoor positioning applications [1].

In recent years, the development of LED technology has transformed the lighting industry by providing high energy efficiency, long lifespan, and low heat emissions. LEDs are now dominant in most lighting applications and are expected to remain so in the near future [2]. Besides replacing traditional lighting methods, LEDs are also a solution for high-precision indoor

positioning systems. The ability to position based on existing infrastructure makes LEDs a practical and cost-effective choice [3].

In study [4], the authors presented several popular indoor VLP techniques. Traditional methods such as angle of arrival (AoA), received signal strength (RSS), and phase difference of arrival (PDoA) serve as the foundation for many systems. However, in recent years, machine learning methods have attracted researchers' attention for improving localization accuracy [5]. Some popular algorithms include decision trees (DT), support vector machines (SVM), artificial neural networks (ANN), reinforcement learning, and deep learning models. Currently, combining several traditional localization techniques with machine learning models has improved localization accuracy and enhanced the stability of indoor visible-light localization systems. This method predicts indoor location with high accuracy by applying the AoA of light and a multi-layer perceptron (MLP) regression model. This proposal achieved a positioning error of 5.2 cm, corresponding to 80% of the measured points, using 4 LEDs and an OpenMV camera [6]. Besides AoA, Time Difference of Arrival (TDoA) has also been applied to enhance indoor VLP by narrowing the fingerprint matching range [7, 8]. The authors in [7] proposed method significantly reduces the matching workload and improves positioning accuracy by combining TDoA with a fingerprint database and using an improved version of K-nearest neighbor (KNN), namely weighted kNN. In [8], N. Chaudhary et al. combined different localization techniques by integrating RSSI-based trilateration with the TDoA method to improve indoor positioning accuracy in VLP systems. The authors in [9] introduce a hybrid indoor VLP system that integrates RSS and TDoA, applying RSS in areas with sufficient signal strength and TDoA elsewhere, resulting in a reduced root-mean-squared error (RMSE) from 7.34 cm to 5.81 cm and improved efficiency. Similarly, the study [10] proposes a quasi-synchronous VLP system combining TDoA and RSS, showing that a two-step method can approach the accuracy of direct positioning at high signal-to-noise ratios. To enhance system robustness under limited LED coverage, the authors in [11] propose a loosely-coupled fusion of a VLP with an Inertial Measurement Unit (IMU) and a rolling shutter camera. Experimental results proved that their proposed method achieved an average accuracy of 2.1 cm, even with the LEDs off and with background light interference. Under limited lighting conditions, the study [12] successfully optimized power distribution among the LEDs to improve positioning accuracy.

In practice, achieving high positioning accuracy is very difficult when the system is affected by environmental noise and transceiver noise. Research [13] proposed a method to measure and mitigate noise using Allan variance for noise analysis and positioning algorithms such as ALSQ and EKF. The results showed that this solution improved by up to 60.4% with an average accuracy of 153 mm. Similarly, to improve the impact of ambient light on the positioning system accuracy, research [14] introduced the RatioVLP method. In addition, another work proposes a method based on Frequency Division Multiplexing (FDM) and real-time signal-to-noise ratio estimation to compensate for positioning errors caused by strong sunlight and other optical conditions. This solution demonstrates that positioning accuracy can be improved using low-cost hardware and a software-defined method [15]. Based on these studies, the authors in [16] examined the impact of signal-dependent shot noise (SDSN) on the visible light communication system. The results showed that SDSN degraded localization performance across all cases, with varying degrees of severity.

In addition to the aforementioned noise sources, the tilt of the LED is also considered a significant factor affecting the accuracy of the VLP. In [17], the authors show that even small tilts of the LED can lead to significant positioning errors. This can result in errors of 1 to 6 cm linearly with the degree of tilt. To overcome the positioning error due to camera tilt in VLP systems, a neural network is used to extract LED image features to correct the tilt. The authors then apply triangulation to perform coordinate estimation [18]. Similarly, the authors in [19] demonstrate that intentionally tilting the transmitter towards a central point on the receiving plane can significantly improve positioning accuracy. Experiments show that the accuracy improves by up to 66% while maintaining uniformity standards of light.

Recently, a novel direction using the Transformer model has been introduced for RSS-based indoor VLP. By effectively extracting hidden spatial features, the approach achieves centimeter-level accuracy. It also maintains stable performance at higher receiving planes, outperforming conventional RSS-based methods [20].

In this study, we focus on improving localization accuracy in L-shaped and T-shaped corridors. These are familiar but particularly challenging environments for localization tasks. These narrow spaces often have limited lighting infrastructure. LEDs are typically installed along a vertical axis and are not symmetrically distributed as in traditional configurations. This reduces the line-of-sight (LOS) signal and increases the presence of the non-line-of-sight (NLOS) signals. Such conditions pose significant difficulties in reducing localization errors. Therefore, we attempt to address these limitations by optimizing signal processing techniques and tuning localization algorithms. The goal of the solution is to enable the system to operate reliably in low-light conditions and in narrow spaces. To evaluate the capabilities of VLP systems in such environments, we propose a novel localization algorithm specifically designed for both LOS and NLOS environments. This study investigates the problem of LED-based indoor localization in complex multipath reflection environments. Specifically, we focus on narrow L-shaped and T-shaped corridors. In these spaces, the LOS assumptions are often weaker, and signal dis-

tortion is severe. A context-aware localization method that combines GraphSAGE and Transformer is proposed. First, noisy RSS measurements are filtered using a global reliability modeling strategy to improve data quality. Next, the cleaned features are then processed by GraphSAGE to capture local spatial relationships. Then, we use a Transformer to model global interactions among RSS signals. To further improve robustness in NLOS conditions, Huber loss is applied to reduce the impact of strong outliers. Finally, the proposed method is evaluated against standalone Transformer, GraphSAGE, Random Forest (RF), and kNN models.

The remainder of this paper is organized as follows. Section 2 presents the layout of the experimental environment and the LED-based localization module. Section 3 introduces the proposed algorithm. Section 4 discusses the experimental setup, evaluation metrics, and comparison. Finally, Section 5 concludes the paper.

2. PHYSICAL SPACE AND LED LOCALIZATION MODULE

In this study, we used two narrow corridors, L-shaped and T-shaped, to establish the experimental space as described in Figs. 1 and 3. The L-shaped corridor is 6 m long and 3 m wide. The T-shaped corridor is 6 m long and 4.6 m wide. In both cases, the LEDs are arranged in the middle and evenly spaced. The lights at the beginning and end of the corridor are 0.6 m from the walls. These arrangements ensure uniform illumination throughout the corridor. Simultaneously, this configuration allows for reliable RSS data collection. As a result, the system can achieve high localization accuracy in environments with limited space and lighting conditions.

To better illustrate the arrangement of the LEDs and photodetectors (PD), Figs. 2 and 4 depict a 3D model of the proposed positioning system. Specifically, ceiling-mounted LEDs are installed along the corridor and act as transmitters in the VLP system. Each LED is installed at a height of 2.8 meters above the floor. The PD is placed on the floor. These 3D images clearly illustrate the relative positions of the LEDs and PD. From this, we can easily visualize the complexity of the research model and the limitations of the light signal in a narrow environment. This configuration closely resembles a real-world indoor environment, with narrow corridors and a specific layout.

In the simulation model, the transmitter units are implemented using ceiling-mounted LEDs. These LEDs serve as the optical signal sources for the positioning system. Each LED emits light following a Lambertian radiation pattern with an emission order of approximately 1.31. It has a semi-angle at half power of 70°. The optical power transmitted by each LED is 500 mW. The PD has a physical detection area of 1 cm^2, which is typical for indoor optical receivers. The PD is assumed to have a field of view (FOV) of 70°. The PD is also equipped with an optical concentrator characterized by a refractive index of 1.5.

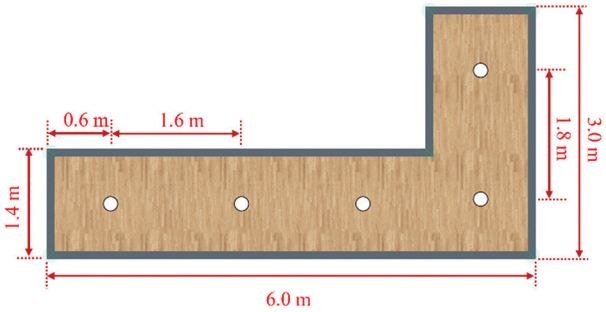


Fig. 1. 2D layout of the L-shaped hallway and LED placement

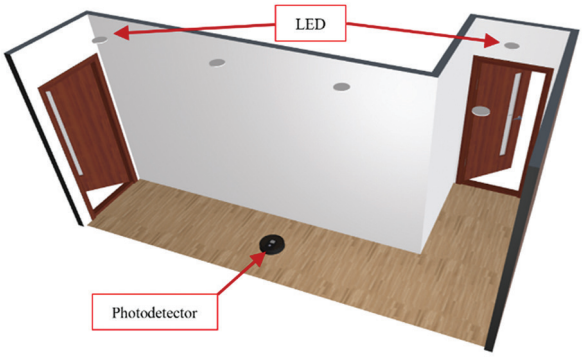


Fig. 2. 3D visualization of the L-shaped hallway setup

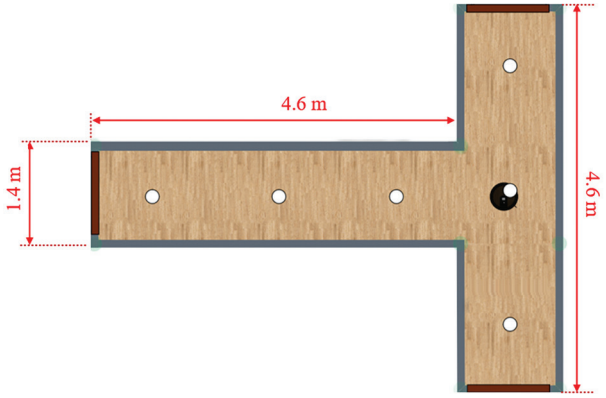


Fig. 3. 2D layout of the T-shaped hallway and LED placement



Fig. 4. 3D visualization of the T-shaped hallway setup

In this paper, we consider the combined effects of shot noise, thermal noise variance, and first-order multipath reflections encountered in practical environments. In VLC systems, NLOS signal propagation arises when light emitted from the LED source reaches the PD after reflecting off intermediate surfaces. The received optical power NLOS depends on several key physical and geometrical factors. It begins with the radiated power from the LED, which follows a Lambertian emission pattern and is given by:

$$P_0 = P_s \frac{(m+1)}{2\pi} \cos^m(\theta) \tag{1}$$

where P_s is the source power, m is the Lambertian order, and θ is the angle of irradiance relative to the surface normal. This emitted light travels through an indirect path involving a reflection. It experiences geometric attenuation over distance, captured by the path loss term:

$$L_d = \frac{1}{d_1^2 d_2^2} \tag{2}$$

with d_1^2 and d_2^2 denoting the distances from the LED to the reflection point and from the reflection point to the PD, respectively.

The interaction with the reflecting surface contributes additional attenuation and angular dependence as:

$$R = \rho \cos(\alpha)\cos(\beta)A_s \tag{3}$$

where ρ is the surface reflection coefficient, α and β are the angles of incidence and reflection, and A_s is the effective reflecting area. The received power is further modulated by the receiver's orientation and optical components, represented by:

$$F(\vartheta_r) = A_{rx} \cos(\vartheta_r) T_f(\vartheta_r) G(\vartheta_r) \tag{4}$$

where ϑ_r is the angle of incidence at the receiver, A_{rx} is the PD area, and $T_f(\vartheta_r), G(\vartheta_r)$ are the optical filter and concentrator gain functions, respectively.

By combining these effects—radiation, distance-dependent loss, surface reflection, and receiver response—the NLOS received power is formulated as:

$$P_{NLOS} = P_0 L_d R F(\vartheta_r) \tag{5}$$

which comprehensively describes the indirect optical power transfer in reflective environments.

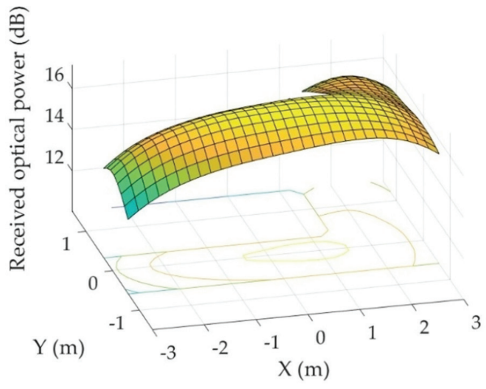


Fig. 5. LOS received optical power in an L-shaped hallway

To visualize how optical signals propagate in the experimental environment, we analyzed the distribution of received optical power under different conditions. Fig. 5 shows the received power with direct LOS from the LEDs. Conversely, Fig. 6 illustrates the power due to reflection from surrounding walls. Although the reflected signal is weaker, it still affects the signal in areas with limited LOS coverage. Finally, Fig. 7 combines both LOS and NLOS components. It provides a more realistic view of the total received power that is ultimately used for positioning.

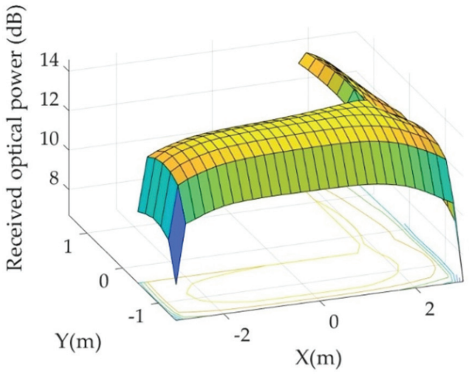


Fig. 6. Received optical power from first-order reflections in and L-shaped hallway

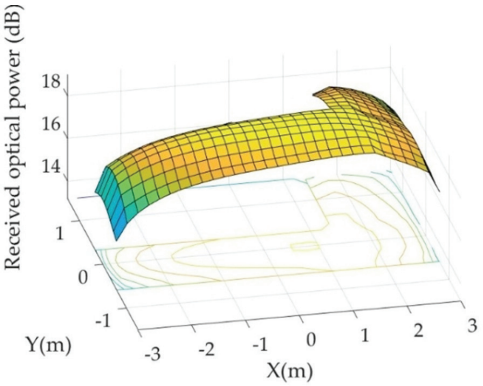


Fig. 7. Total received optical power in an L-shaped hallway

3. PROPOSED CONTEXT-AWARE GRAPHSAGE - TRANSFORMER ALGORITHM

To effectively estimate indoor positions from received optical signals, we propose an improved deep learning architecture, Context-Aware GraphSAGE–Transformer (CA-GST), which combines context-aware GraphSAGE with a Transformer encoder.

Graph Sample and Aggregate (GraphSAGE) is a framework designed for inductive representation learning on large graphs [21]. Instead of using the full graph structure, GraphSAGE samples a fixed-size neighborhood and learns to aggregate information from neighboring nodes. This allows it to generate node embeddings for unseen data and scale effectively to large, dynamic graphs.

Transformer, originally developed for natural language processing tasks, is based on a self-attention mechanism that allows it to capture relationships between all elements in a sequence regardless of their positions [22]. It has been widely adopted in various fields, including computer vision and graph learning, due to its ability to model long-range dependencies and its highly parallelizable architecture.

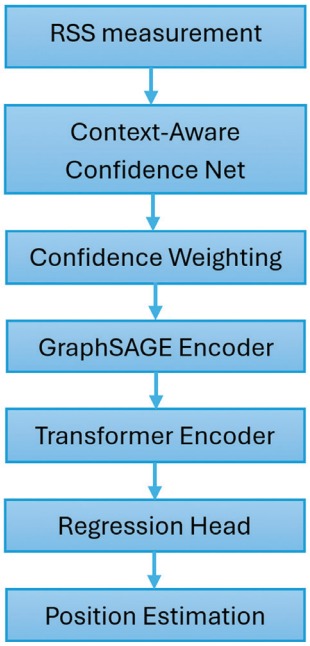


Fig. 8. Block diagram of proposed solution

In this paper, we propose a context-aware GraphSAGE–Transformer model for robust indoor localization under NLOS conditions, as shown in Fig. 8. The model first introduces an early-stage context-aware reliability modeling strategy that learns global RSS consistency patterns across multiple LEDs. By adaptively reweighting LED signals before graph-based message passing, unreliable RSS measurements are suppressed, preventing noise propagation in subsequent feature aggregation. Next, GraphSAGE is employed to model local spatial relationships among LEDs, where each LED is treated as a graph node and neighborhood aggregation captures inter-LED signal correlations induced by spatial proximity and shared propagation paths. Then, a Transformer encoder is applied to the graph-enhanced features to capture global and long-range dependencies across all LEDs. Through self-attention, the Transformer dynamically assigns importance to different LEDs based on their global relevance. Finally, the Huber loss is adopted during training to improve robustness against severe NLOS-induced outliers further. Because Huber loss combines the advantages of both MSE and MAE, it behaves quadratically for small residuals and linearly for large ones.

The model configuration details are summarized in Table 1, where the input includes RSS vectors, and the output consists of the predicted 2D coordinates along with the localization error metric. All configuration details of the proposed model, including network architecture parameters, GraphSAGE and Transformer settings, context-aware module design, training hyperparameters, and optimization strategy, are also summarized in this table.

Table 1. Model configuration

Type	Component	Description
Input	RSS Measurements	$X \in R^{N\times M}$: RSS from M LEDs at N locations. For an L-shaped hallway, $M = 5$ and $N = 320$; for a T-shaped hallway, $M = 6$ and $N = 392$.
	Coordinates	$Y \in R^{N\times 2}$: Ground-truth 2D positions (x, y)
	Graph Structure	$G = (V, E)$: Grid graph connecting spatial neighbors
Main process	Train-test split	80% training samples and 20% testing samples
	GraphSAGE layers	2
	Layer type	SAGEConv
	Initial hidden dimension	32
	Embedding dimension	64
	Activation function	ReLU
	Transformer type	Encoder
	Encoder layer	2
	Number of attention heads	4
	Optimizer	Adam
	Learning rate	1×10^{-3}
	Epochs	300
	Loss function	Huber
	Huber threshold	1.0
Output	Predicted Positions	$\hat{Y} \in R^{N\times 2}$: Estimated positions output by the model
	Evaluation Metric	RMSE, MAE and ME between $\hat{Y}$ and Y on test samples.

The detailed procedure of the proposed CA-GST algorithm is described in Table 2.

Table 2. CA-GST Algorithm

Input:
• RSS fingerprint dataset X
• Ground-truth positions Y
• AP graph topology $G = (V, E)$
• Maximum training epochs E
• Learning rate η
• Huber loss threshold δ
Process:
• Normalize RSS fingerprints X and position labels Y
• Initialize: – Context-aware confidence network – GraphSAGE layers – Transformer encoder – Regression output layer

Repeat for epoch = 1 to E

For i = 1 to N (training samples)

- Estimate context-aware confidence weights c_i from RSS fingerprint x_i
- Reweight RSS features using element-wise multiplication $x_i \leftarrow x_i \odot c_i$
- Project each RSS value into the latent feature space
- Perform GraphSAGE neighborhood aggregation using graph G
- Apply the Transformer encoder to capture global inter-AP dependencies
- Predict user position $\hat{Y}_i$ from aggregated features
- Compute localization loss using Huber loss between $\hat{Y}_i$ and Y_i
- Update all model parameters via backpropagation

End for

End Repeat

Output: Estimated positions $\hat{Y}$ for all RSS fingerprints

Based on the detailed algorithm description in Table 2, the specific steps of the training process and results evaluation are as follows:

- Data Normalization

Normalize input RSS values (X) and target coordinates (Y) using standard score normalization:

$$x_{scaled} = \frac{X - \mu_x}{\sigma_x}, y_{scaled} = \frac{Y - \mu_y}{\sigma_y} \tag{6}$$

where μ_x, σ_x are the mean and standard deviation of X, similarly for Y.

- Graph Construction

An undirected graph $G = (V, E)$ is constructed to model inter-LED signal relationships. Each node $v_i \in V$ represents one LED, and edges connect adjacent LEDs:

$$E = \{(v_i, v_{i+1}), (v_{i+1}, v_i)\}, \quad i = 1, \ldots, M-1 \tag{7}$$

This graph structure enables GraphSAGE to aggregate contextual information across neighboring LEDs.

- Context-Aware Feature Reweighting

Before graph-based feature extraction, a context-aware mechanism is introduced to reweight the RSS inputs adaptively. Given an RSS vector $x_i \in R^M$, a lightweight gating network produces a confidence vector:

$$c_i = \sigma(MLP(x_i)) \in R^M \tag{8}$$

The reweighted input is then computed using residual guidance:

$$\tilde{x}_i = x_i \odot c_i + \tau x_i \tag{9}$$

where $\odot$ denotes element-wise multiplication and τ is a residual coefficient.

- Feature Extraction with GraphSAGE

To capture contextual dependencies among the LED signals, each RSS vector $x_i \in R^M$ is first reshaped into a column vector:

$$x_i^{reshaped} = reshaped(x_i) \in R^{M \times 1} \tag{10}$$

It is then projected into a higher-dimensional space using a linear transformation:

$$z_i = W_{proj} x_i^{reshaped} + b_{proj}, \; z_i \in R^{M \times d_s} \tag{11}$$

where $W_{proj} \in R^{d_s \times 1}$ and $b_{proj} \in R^{d_s}$ are learnable parameters. d_s is the initial hidden dimension of GraphSAGE.

Then, two layers of GraphSAGE aggregate features from neighbor LEDs:

$$h_i^{(1)} = \sigma(AGGREGATE^{(1)}(z_i, N(i))) \tag{12}$$

$$h_i^{(2)} = \sigma(AGGREGATE^{(2)}(h_i^{(1)}, N(i))), \quad h_i^{(1)}, h_i^{(2)} \in R^{d_m} \tag{13}$$

where:

- $N(i)$ denotes the neighbors of node i
- σ is a non-linear activation
- $AGGREGATE^{(n)}$ is the GraphSAGE aggregation function
- d_m is embedding dimension

- Sequence Modeling with Transformer Encoder

The set of LED feature vectors after GraphSAGE is treated as:

$$h_i = [h_1^{(2)}, h_2^{(2)}, \ldots, h_5^{(2)}] \in R^{M \times d_m} \tag{14}$$

Learnable LED embeddings $E_{pos} \in R^{5 \times d_m}$ are added to h_i to serve as positional encodings:

$$\tilde{h}_i = h_i + E_{pos} \tag{15}$$

The resulting sequence is passed through a multi-layer Transformer encoder to extract latent features:

$$H_i = TransformerEncoder(\tilde{h}_i) \in R^{M \times d_m} \tag{16}$$

Finally, the encoder output is flattened to yield a fixed-size feature vector:

$$f_i = Flatten(H_i) \in R^{M \times d_m} \tag{17}$$

- Position Prediction Layer

Finally, the model predicts the 2D coordinates via a fully connected regression head:

$$\hat{Y}_i = W_{out} f_i + b_{out}, \; \hat{Y}_i \in R^2 \tag{18}$$

here, $\hat{Y}_i$ denotes the predicted 2D position for the i^{th} sample; W_{out} is the trainable weight matrix of the regression layer; b_{out} is the bias.

- Training Loss

The model is trained by minimizing the Huber loss between the predicted coordinates $\hat{Y}_i$ and the ground-truth coordinates Y_i:

$$L = \begin{cases} \frac{1}{2}(Y - \hat{Y})^2 & , |Y - \hat{Y}| \le \delta \\ \delta |Y - \hat{Y}| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases} \tag{19}$$

where δ is the transition threshold.

- Evaluation Metrics

Performance is assessed on a separate test set of M samples. Three standard regression metrics are used:

$$RMSE = \sqrt{(\frac{1}{M}\sum_{j=1}^{M}\|\hat{Y}_j - Y_j\|^2)} \tag{20}$$

$$MAE = \frac{1}{M}\sum_{j=1}^{M}\|\hat{Y}_j - Y_j\|_1 \tag{21}$$

$$ME = \frac{1}{M}\sum_{j=1}^{M}(\hat{Y}_j - Y_j) \tag{22}$$

These metrics quantify the Euclidean and Manhattan distances between predicted and true positions, respectively.

4. EXPERIMENTAL EVALUATION AND COMPARATIVE ANALYSIS

In this study, we conducted a comprehensive experimental evaluation to compare the localization accuracy of five models: CA-GST, Transformer, GraphSAGE, RF, and kNN. As illustrated in Fig. 9, the performance of each model was assessed using standard statistical metrics, including RMSE, MAE, and ME.

For both hallway configurations—L-shaped and T-shaped—the CA-GST model delivered the strongest results, achieving RMSE, MAE, and ME of 0.20 m, 0.13 m, and 0.17 m, respectively. This demonstrates the effectiveness of combining the topological information learned by GraphSAGE with the Transformer's sequential modeling power. In addition, the context-aware feature reweighting mechanism adaptively suppresses unreliable RSS measurements caused by NLOS effects. At the same time, the use of Huber loss improves robustness against severe outliers during training. When combined, these components help the model capture both spatial relationships and broader signal interactions, which in turn improve localization accuracy.

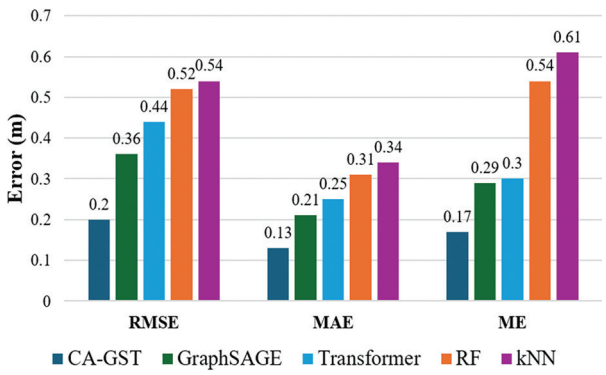


Fig. 9. Error comparison

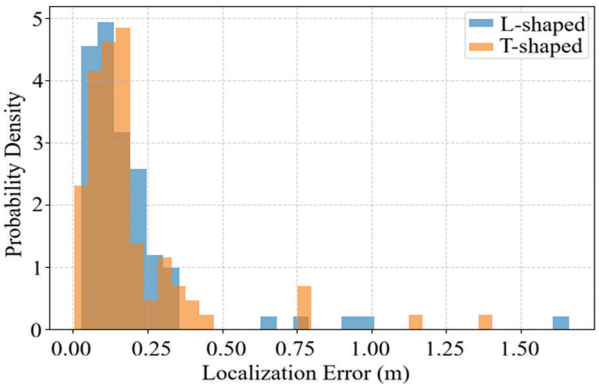


Fig. 10. Histogram of localization error

In contrast, the Transformer model recorded an RMSE of 0.44 m, an MAE of 0.25 m, and a Median Error of 0.30 m. Despite being capable of modeling sequence patterns, its lack of structural awareness in the spatial context results in higher localization errors. The GraphSAGE model performed slightly better than the Transformer regarding RMSE (0.36 m) and MAE (0.21 m), but still underperformed the complete hybrid model. Furthermore, to gain a more comprehensive understanding of the effectiveness of the proposed algorithm, we also compared it with popular machine learning models such as RF and kNN. Experimental results showed that the CA-GST model has significantly lower errors than the machine learning models. Specifically, the RF and KNN algorithms have an RMSE approximately three times higher than our solution. These results indicate that the proposed CA-GST model significantly improves localization accuracy over the individual models.

To better understand the model's performance, the histogram and cumulative distribution function (CDF) illustrate its localization accuracy as depicted in Figs. 10 and 11. The histogram in both cases, with the L-shaped and T-shaped hallways, shows that most errors are below 0.3 meters, indicating high precision with only a few outliers above 0.5 meters.

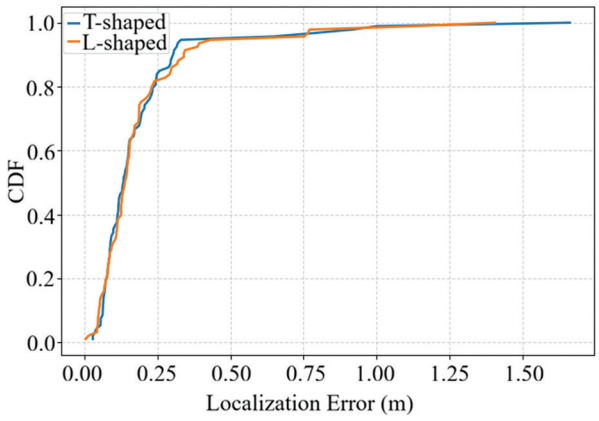


Fig. 11. CDF of localization error

The CDF confirms this trend, with 80% of predictions under 0.25 meters and nearly all within 0.35 meters. These results highlight the stability and reliability of the CA-GST

model, even in complex indoor environments. Finally, to provide a more intuitive evaluation of the coordinate estimation results from the CA-GST algorithm, we visualize the absolute error distribution on both the X and Y axes, as shown in Fig. 12. This plot offers a clearer comparison of prediction accuracy across different models or configurations by highlighting the extent to which the estimated positions deviate from the ground truth along each axis.

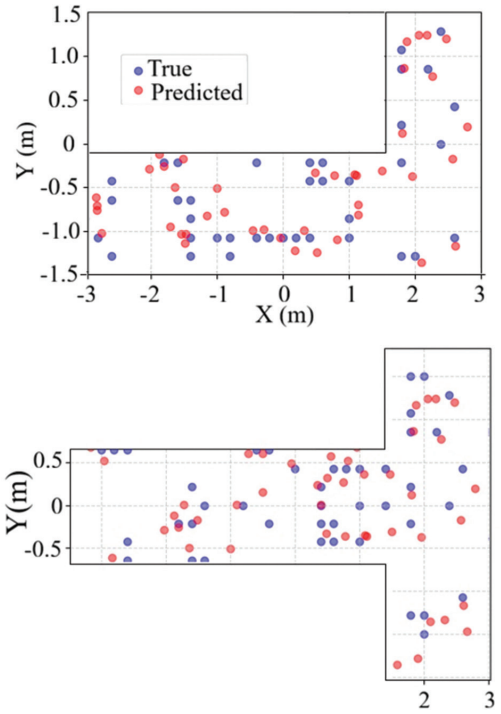


Fig. 12. Heatmap of predicted vs. true positions

Table 3. Localization performance of the CA-GST algorithm using 5-fold cross-validation

Hallway Type	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	RMSE (m)	Deviation from Mean (m)
L-shaped	Test	Train	Train	Train	Train	0.215104	+0.008881
	Train	Test	Train	Train	Train	0.172523	-0.033700
	Train	Train	Test	Train	Train	0.210148	+0.003925
	Train	Train	Train	Test	Train	0.202037	-0.004186
	Train	Train	Train	Train	Test	0.231302	+0.025079
T-shaped	Test	Train	Train	Train	Train	0.221113	+0.011309
	Train	Test	Train	Train	Train	0.184532	-0.025272
	Train	Train	Test	Train	Train	0.192331	-0.017473
	Train	Train	Train	Test	Train	0.238721	+0.028917
	Train	Train	Train	Train	Test	0.212321	+0.002517

To evaluate the stability of the proposed algorithm on the system's dataset, cross-validation is applied. This is an important technique for assessing the model's generalizability to data that has never appeared during training. Instead of dividing the data into only two sets, training and testing, we divided the dataset into five groups. The model is trained multiple times, each time using a portion as the test set and the remaining portions as the training set. The experimental results are detailed in Table 3. The proposed CA-GST algorithm achieved an average RMSE of 0.2062 ± 0.0194 m and 0.2098 ± 0.0212 m in the L-shaped and T-shaped hallways, respectively, after 5-fold cross-validation. The results show that the RMSE value remained stable across 5 assessments for both types of hallways.

5. CONCLUSION

In this paper, a context-aware GraphSAGE–Transformer model is proposed, which integrates reliability modeling to capture both local spatial relationships and global signal dependencies. The proposed method is designed for challenging environments such as L-shaped and T-shaped corridors, where NLOS conditions and limited lighting infrastructure significantly affect localization performance. This study demonstrated the effectiveness of our proposed indoor localization algorithms in narrow hallway environments with limited LED infrastructure. The proposed CA-GST method helped achieve a root mean squared error (RMSE) of 0.2 meters, representing a 55% reduction compared to the Transformer model and a 44% reduction compared to GraphSAGE alone. Furthermore, the proposed algorithm has an RMSE approximately three times lower than that of traditional machine learning algorithms, such as RF and kNN. These improvements are especially significant given the 1.4-meter-wide, oddly shaped hallway and the limitations imposed by the number of available LED sources.

However, the proposed method still has some limitations. It requires further validation in larger environments with more obstacles. In addition, the method has relatively high computational cost, which may affect real-time performance.

Future work will focus on deploying this model in real-time systems and evaluating its performance on mobile indoor robots to enable reliable navigation in complex, constrained indoor spaces.

DECLARATION OF THE USE OF AI TOOLS

AI tools were not used.

ACKNOWLEDGMENTS

We would like to thank Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam for the support this study.

REFERENCES

[1] J. Armstrong, Y. A. Sekercioglu, A. Neild, "Visible Light Positioning: A Roadmap for International Standardization", IEEE Communications Magazine, Vol. 51, No. 12, 2013, pp. 68-73. doi: 10.1109/MCOM.2013.6685759

[2] S. Bastiaens, M. Alijani, W. Joseph, D. Plets, "Visible Light Positioning as a Next-Generation Indoor Positioning Technology: A Tutorial", IEEE Communications Surveys & Tutorials, Vol. 26, No. 4, 2024, pp. 2867-2913. doi: 10.1109/COMST.2024.3372153

[3] Z. Zhu, Y. Yang, M. Chen, C. Guo, J. Cheng, S. Cui, "A Survey on Indoor Visible Light Positioning Systems: Fundamentals, Applications, and Challenges", IEEE Communications Surveys & Tutorials, Vol. 27, No. 3, 2025, pp. 1656-1686. doi: 10.1109/COMST.2024.3471950

[4] H. Q. Tran, C. Ha, "Machine Learning in Indoor Visible Light Positioning Systems: A Review", Neurocomputing, Vol. 491, 2022, pp. 117-131. doi: 10.1016/j.neucom.2021.10.123

[5] M. H. Moharam, H. A. El-Hakim, "Support Vector Machines for Indoor Visible Light Positioning: A Comparative Study with Neural Networks", Journal of Optics, 2026. doi: 10.1007/s12596-026-03075-5

[6] G. He, Y. Li, B. Bo, Y. Qian, P. Wang, Q. Zhang, "Accuracy Evaluation of Indoor AoA Visible Light Positioning with MLP Regression Model", IEEE Photonics Technology Letters, Vol. 37, No. 11, 2025, pp. 637-640. doi: 10.1109/LPT.2025.3545063

[7] C. Zhang, D. Li, X. Feng, L. Tan, L. Yang, X. Yang, "Indoor Visible Light Positioning Method Based on TDOA and Fingerprint", Proceedings of the IEEE International Conference on Electronic Technology, Communication and Information, Changchun, China, 27-29 August 2021, pp. 397-401. doi: 10.1109/ICETCI53161.2021.9563479

[8] S. M. Sheikh, H. M. Asif, K. Raahemifar, F. Al-Turjman, "Time Difference of Arrival Based Indoor Positioning System Using Visible Light Communication", IEEE Access, Vol. 9, 2021, pp. 52113-52124. doi: 10.1109/ACCESS.2021.3069793

[9] M. Karmy, S. ElSayed, A. Zekry, "Performance Enhancement of an Indoor Localization System Based on Visible Light Communication Using RSSI/TDOA Hybrid Technique", Journal of Communications, Vol. 15, No. 5, 2020, pp. 379-385.

[10] E. Kazikli, S. Gezici, "Hybrid TDOA/RSS Based Localization for Visible Light Systems", arXiv:1802.03991, 2018.

[11] W. Guan, L. Huang, B. Hussain, C. P. Yue, "Robust Robotic Localization Using Visible Light Positioning and Inertial Fusion", IEEE Sensors Journal, Vol. 22, No. 6, 2022, pp. 4882-4892. doi: 10.1109/JSEN.2021.3053342

[12] M. F. Keskin, A. D. Sezer, S. Gezici, "Optimal and Robust Power Allocation for Visible Light Positioning Systems under Illumination Constraints", IEEE Transactions on Communications, Vol. 67, No. 1, 2019, pp. 527-542. doi: 10.1109/TCOMM.2018.2866849

[13] Y. Zhuang, et al., "Visible Light Positioning and Navigation Using Noise Measurement and Mitigation", IEEE Transactions on Vehicular Technology, Vol. 68, No. 11, 2019, pp. 11094-11106. doi: 10.1109/TVT.2019.2943517

[14] X. Yang, Y. Zhuang, M. Shi, X. Sun, X. Cao, B. Zhou, "RatioVLP: Ambient Light Noise Evaluation and Suppression in the Visible Light Positioning System", IEEE Transactions on Mobile Computing, Vol. 23, No. 5, 2024, pp. 5755-5769. doi: 10.1109/TMC.2023.3312550

[15] A. Costanzo, V. Loscri, "Visible Light Indoor Positioning in a Noise-Aware Environment", Proceedings of the IEEE Wireless Communications and Networking Conference, Marrakesh, Morocco, 15-18 April 2019, pp. 1-6. doi: 10.1109/WCNC.2019.8885607

[16] A. Cheema, M. Alsmadi, S. Ikki, "Effect of Signal-Dependent Shot Noise on Visible Light Positioning", IEEE Photonics Journal, Vol. 14, No. 3, 2022, p. 7327907. doi: 10.1109/JPHOT.2022.3170693

[17] D. Plets, S. Bastiaens, L. Martens, W. Joseph, "An Analysis of the Impact of LED Tilt on Visible Light Positioning Accuracy", Electronics, Vol. 8, No. 4, 2019, p. 389. doi: 10.3390/electronics8040389.

[18] T. Yuan, Y. Xu, Y. Wang, P. Han, J. Chen, "A Tilt Receiver Correction Method for Visible Light Positioning Using Machine Learning Method", IEEE Photonics Journal, Vol. 10, No. 6, 2018, p. 7909312. doi: 10.1109/JPHOT.2018.2880872

[19] N. Chaudhary, O. I. Younus, L. N. Alves, Z. Ghassemlooy, S. Zvanovec, H. Le-Minh, "An Indoor Visible Light Positioning System Using Tilted LEDs with High Accuracy", Sensors, Vol. 21, No. 3, 2021, p. 920. doi: 10.3390/s21030920

[20] Y. Yao, Y. Zhu, X. Shao, "Indoor Positioning of Visible Light Communication Based on Transformer Model", Guangtongxin Yanjiu, 2026. doi: 10.13756/j.gtxyj.2026.240201

[21] W. L. Hamilton, R. Ying, J. Leskovec, "Inductive Representation Learning on Large Graphs", Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5-10 December 2016, pp. 1024-1034.

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention Is All You Need", Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5-10 December 2016, pp. 5998-6008.