

Article

Indoor Positioning by Double Deep Q-Network in VLC-Based Empty Office Environment

Sung Hyun Oh  and Jeong Gon Kim * 

Department of Electronic Engineering, Tech University of Korea, Siheung-si 15297, Republic of Korea; osh119@tukorea.ac.kr

* Correspondence: jgkim@tukorea.ac.kr; Tel.: +82-10-9530-5103

Abstract: Recently, artificial intelligence (AI) has been applied in various industries. One such application is indoor user positioning using Big Data. The traditional method for positioning is the global positioning system (GPS). However, the performance of GPS is limited indoors due to propagation loss. Hence, radio frequency (RF)-based communication methods such as WiFi and Bluetooth have been proposed as indoor positioning solutions. However, positioning performance inaccuracies arise due to signal interference caused by RF band saturation. Therefore, this study proposes indoor user positioning based on visible light communication (VLC). The proposed method involves the sequential application of fingerprinting and double deep Q-Network. Fingerprinting is utilized to define the action and state of the double deep Q-Network agent. The agent is designed to learn and locate the reference point (RP) closest to the user's position in a shorter search time. The core idea of the proposed system is to converge a Cell-ID scheme and fingerprinting. Through this, the initial state of the double deep Q-Network agent can be limited. A limited initial state can increase the positioning speed. Simulation results show that the proposed scheme attains a positioning resolution of less than 13 cm and achieves a processing time of less than 0.03 s to obtain the final position in VLC-based office environments.

Keywords: indoor positioning; artificial intelligence (AI); visible light communication (VLC); received signal strength (RSS); deep reinforcement learning (DRL)



Citation: Oh, S.H.; Kim, J.G. Indoor Positioning by Double Deep Q-Network in VLC-Based Empty Office Environment. *Appl. Sci.* **2024**, *14*, 3684. <https://doi.org/10.3390/app14093684>

Academic Editors: Yun Seop Yu, Dongsik Jo, Hee-Cheol Kim, Kwang-Baek Kim and Jongtae Lee

Received: 10 March 2024

Revised: 15 April 2024

Accepted: 22 April 2024

Published: 26 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of artificial intelligence (AI), it can be applied in various industrial fields [1,2]. Recently, research on location-based services (LBSs) has been conducted in various industries to meet individual and public needs [3]. LBSs include a range of applications, including navigation, logistics, emergency services, and security. While outdoor positioning has achieved relatively accurate results through advancements in the global positioning system (GPS) [4], the positioning performance of GPS is significantly diminished indoors due to propagation loss caused by obstacles and walls. As a result, efforts are being made to address this issue and improve the accuracy of indoor user positioning [5].

The aforementioned studies primarily focus on wireless communication technologies and positioning algorithms applicable in indoor environments. Prominent wireless communication technologies include radio frequency (RF)-based Bluetooth, WiFi, and ultrawideband (UWB) [6–8]. However, these technologies may experience degraded positioning performance due to RF band saturation. To solve this problem, visible light communication (VLC)-based positioning is raised in this paper. VLC utilizes light-emitting diodes (LEDs) to provide both lighting and communication capabilities simultaneously [9,10]. LEDs are increasingly replacing indoor lighting fixtures due to their long lifespan and high energy efficiency. Using these characteristics, VLC is considered one of the next-generation mobile communication technologies [11].

Existing indoor positioning algorithms include time of arrival (TOA), angle of arrival (AOA), time difference of arrival (TDOA), and received signal strength (RSS) [12]. Among these, the RSS-based positioning algorithm is widely adopted due to its simplicity of implementation and lack of additional hardware requirements. Although RSS-based positioning has relatively low accuracy, which is attributed to the issue of signal variability, the accuracy can be improved when combined with the fingerprinting technique [13].

Recently, there has been active research on applying AI technology to positioning methods based on RSS data [14]. AI technology enables real-time positioning due to its fast processing speed and can achieve a high-precision positioning performance when provided with sufficient training data [15].

In this study, we propose a technique for user positioning in an indoor VLC environment using a double deep Q-Network. Our focus is on achieving a precise positioning performance while minimizing processing time. To achieve this, we construct a fingerprinting database and train a double deep Q-Network model. Traditional fingerprinting database-based matching techniques suffer from increased processing time as the fingerprint database grows larger. However, our proposed double deep Q-Network model significantly reduces processing time while improving positioning accuracy. Furthermore, in this study, we take into account the impact of multipath reflections on the positioning performance. The main contributions of this paper are described as follows:

- (1) The core technology of this study is the application of the double deep Q-Network to significantly improve the processing speed of the fingerprinting technique.
- (2) The Cell-ID technique can be used to speed up convergence by selecting the initial location of the reinforcement learning agent.
- (3) Through simulation results, it can be seen that a fast processing time is achieved when the initial location of the agent is selected.

The paper is structured as follows. Section 2 analyzes related work for performing indoor positioning. In Section 3, the system model used in this study is described. Section 4 presents the proposed indoor positioning method. In Section 5, the simulation parameters and results are presented. In Section 6, the conclusion is summarized, and future work directions are explained in view of improving indoor positioning.

2. Related Work

In this section, we analyze related work that performed positioning based on AI in an indoor VLC environment. AI can be broadly divided into supervised learning, unsupervised learning, and reinforcement learning. Therefore, in this section, we analyze the positioning using each learning method.

First, we analyze the related literature that performed supervised learning for positioning. In [16], the authors introduced a method for multiple fingerprint positioning based on artificial neural networks (ANNs). They achieved a positioning performance of 2.68 m while simultaneously positioning 15 targets, demonstrating high accuracy even in environments with radio interference. However, they encountered an issue when a significant error of 2.29 m occurred even with a single target. In [17], the authors propose a method to estimate the location of a moving user in an indoor VLC environment. The authors applied the time division multiple access (TDMA) to eliminate inter-signal interference. The proposed positioning method uses the received power from three transmitters as the input for the ANN and the distance between the transmitter and receiver as the output. Simulation results showed that the neural network estimation approach provides more accurate positioning than the trilateration estimation with 94% accuracy when the receiver's normal has a tilt angle with the transmitter's normal.

Second, we analyze the related literature that performed unsupervised-learning-based positioning. In [18], the authors proposed a method that utilizes the weighted k-nearest neighbor (kNN), an unsupervised learning method within the field of AI technologies. Their approach achieved a more precise positioning performance compared to conventional triangulation methods. In [19], the authors combined the unsupervised algorithm kNN

and traditional machine learning random forest. In the results, a precise positioning performance of 0.069 m at the center, 0.157 m at the edge, and 0.166 m at the corner was found using their proposed method. In [20], the authors propose the adaptive residual weighted k-nearest neighbor (ARWKNN) technique to solve the k-value optimization problem to obtain the minimum positioning error in WKNN. The proposed method adaptively selects the k-value by matching the residual between the measured and calculated RSSI values. In other words, the k-value corresponding to the smallest residual sum of squares of the RSSI is obtained. Through simulation, the proposed ARWKNN significantly reduced the average position error while maintaining a similar level of complexity compared to the existing WKNN and KNN.

Third, we analyze the related literature that performed reinforcement-learning-based positioning. In [21], the authors demonstrated a three-dimensional indoor VLC positioning system using multiple photodiodes and reinforcement learning to enhance positioning accuracy. To achieve this goal, the authors created a VLC positioning system using multiple photodiodes and enhanced performance through the application of a reinforcement learning algorithm. The experimental results demonstrate that VLC using multiple photodiodes can effectively determine location in a three-dimensional space. The results also show improved accuracy and performance through reinforcement learning. In [22], the primary objective of the research is to achieve precise positioning using visible light. To accomplish this, point-wise reinforcement learning is used iteratively. This involves maximizing performance through iterative improvement in the process of learning and applying strategies to enhance positioning accuracy. Simulation results confirm that the proposed IPWRL achieves a lower position error compared to the existing RSS algorithm and PWRL.

Therefore, in this work, we try to use a fingerprinting scheme [23], Cell-ID scheme [24], and reinforcement learning [20,21] to improve the positioning accuracy. The authors of [23] proposed a localization technique using fingerprinting and KNN. A database of RPs is constructed using fingerprinting, and nearest neighbor RPs are obtained through an iterative search process. Afterward, the position of the UE is determined using a weighted average. In [23], the location of the UE could be determined by sequentially applying simple techniques. As the number of reference points increases, the time for repeated searches also increases significantly. The authors of [24] performed transmitter identification and UE position determination using RGB LEDs. The transmitter transmits an LED ID using frequency division multiplexing (FDM), and the UE receives it to identify the location of the transmitter. Afterward, the location of the UE is determined using a triangulation technique. In [24], the authors achieved high positioning accuracy using RGB LEDs. However, there is a limitation that it can only be applied to indoor environments using the CIE 1931 standard.

This study improved the performance of the double deep Q-Network model by applying fingerprinting and Cell-ID techniques based on an analysis of the existing literature. In the case of reinforcement learning, there is an issue with the initial state being set randomly, which leads to prolonged optimization processes. Therefore, the time for optimization can be reduced by limiting the initial state of the agent using the Cell-ID method. In addition, quick optimization was performed using the learned agent without searching all reference points in the fingerprint database. Detailed information regarding this is covered in Section 4.

3. System Model

In this section, the system model used to evaluate positioning performance will be described in detail. First, the indoor environment considered in this study is shown in Figure 1. The size of the indoor environment is 5 m × 5 m × 3 m, and it is assumed to be an empty space.

In this environment, it is assumed that a total of four LED access points (APs) are placed at a height of 3 m, and the receiver moves parallel to the ground at a height of 0.7 m. The transceiver characteristics are summarized in Table 1.

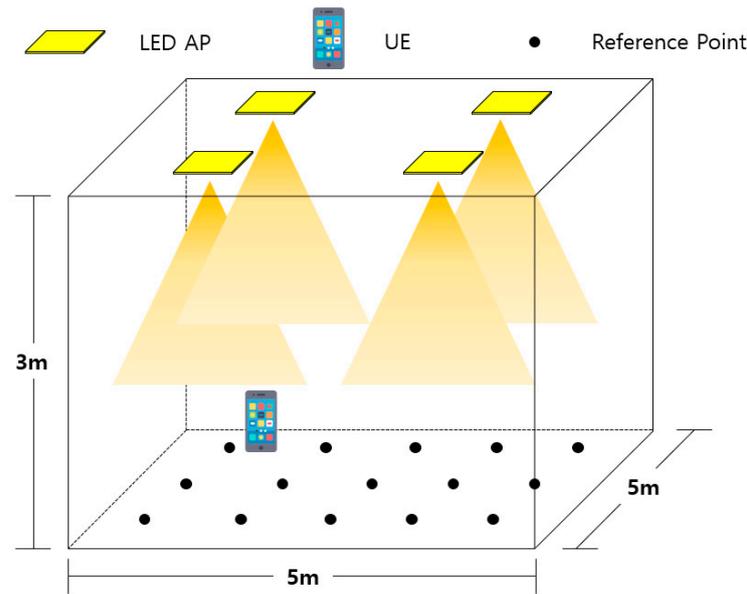


Figure 1. Indoor VLC system model.

Table 1. Transceiver characteristics.

	Parameter	Value
Transmitters	Transmit power	10 W
	Half power semi-angle	60°
	Elevation	−90°
	Field of view (FOV)	60°
Receiver	Active area	1 cm ²
	Gain of optical filter	1
	Gain of optical concentrator	1.5

Next, we analyze the channel characteristics of the line of sight (LOS) and non-line of sight (NLOS) paths in an indoor environment. The path of visible light in the indoor VLC environment is illustrated in Figure 2. The path of visible light can be divided into two components. The green line represents the LOS path while the red and yellow lines represent the NLOS path. The NLOS path can be further subdivided into an incident wave from the LED AP to the wall (red line) and a reflected wave from the wall to the receiver (yellow line). The channel characteristics associated with each path are as follows:

$$h_{u,dir}^i = P_t \frac{A(m+1)}{2\pi d^2} \cos^m(\theta) T_s(\psi) C(\psi) \cos(\psi) \tag{1}$$

$$h_{u,non-dir}^i = P_t \frac{A(m+1)}{2\pi d_1^2 d_2^2} \rho \cos^m(\theta) dA_{wall} \cos(\alpha) \cos(\beta) T_s(\psi_r) C(\psi_r) \cos(\psi_r) \tag{2}$$

where (1) and (2) are the RSS according to the LOS and NLOS paths, respectively. First, $h_{u,dir}^i$ and $h_{u,non-dir}^i$ denote the RSS of each path and P_t denotes the LED AP transmit power. A is the active area of the receiver and m is the Lambertian order. d is the distance from the LED AP to the receiver. d_1 is the distance from the LED AP to the reflective surface, and d_2 is the distance from the reflective surface to the receiver. ρ is the reflection coefficient of the wall and dA_{wall} is the surface element of the wall. Also, $T_s(\psi)$ and $C(\psi)$ denote the optical filter gain and optical concentration gain, respectively, and ψ_c means the FOV of the receiver.

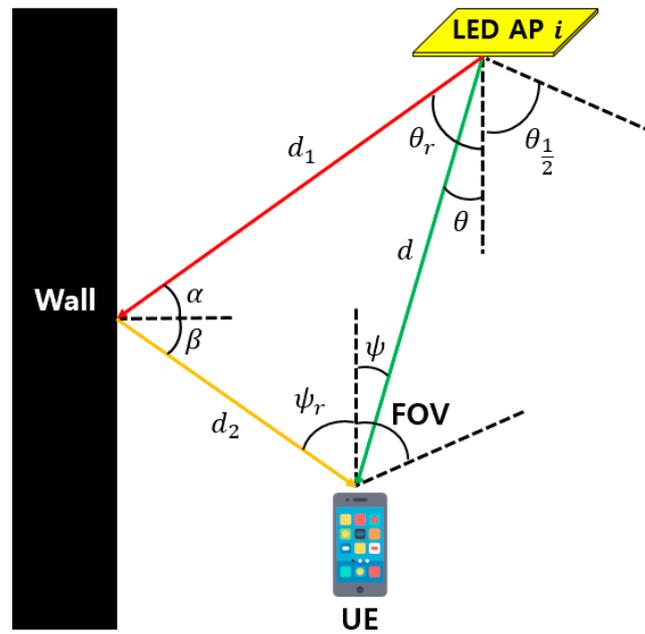


Figure 2. Indoor VLC scenario.

Based on the above Equations (1) and (2), the overall RSS distribution of the indoor environment applied in this paper is shown in Figure 3.

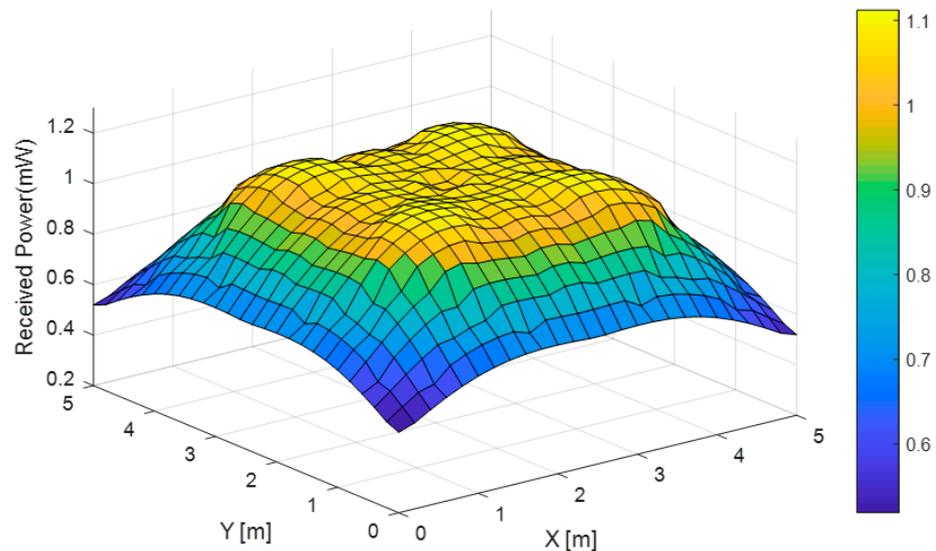


Figure 3. Overall RSS distribution.

4. Proposed Indoor Positioning Method

In this section, we describe the proposed positioning method that combines the fingerprinting technique with the double deep Q-Network. The key components of the proposed method are the construction of an RSS database using fingerprinting and the training of the DDQN model. Additionally, a block diagram of the proposed method is presented at the end of this section.

Initially, the fingerprinting technique is employed to construct an RSS database specific to the indoor environment. Subsequently, the double deep Q-Network model is trained. To facilitate this, the elements of reinforcement learning—the agent, state, action, and reward—are defined. Once the model training is complete, the performance of the positioning is evaluated using the trained model.

4.1. Fingerprinting

The fingerprinting technique is a positioning method that can achieve a precise positioning performance when combined with the RSS. It involves designating specific locations, called reference points (RPs), within the indoor environment. By increasing the number of RPs and making them denser, more accurate positioning can be achieved. However, as the number of RPs increases, the processing time also increases. Hence, it is crucial to determine the optimal number of RPs, taking into account the size of the indoor environment.

The fingerprinting technique consists of two steps: offline and online. The first step is the offline phase, where the RSS database is constructed for each reference point without a specific positioning target. Typically, RP arrangement and RSS correction are performed in this step. The RSS database for the r -th reference point is stored as shown in (3):

$$RSS_{DB} = \begin{bmatrix} h_1^1 & \cdots & h_1^i & \cdots & h_1^L \\ \vdots & & \vdots & & \vdots \\ h_r^1 & \cdots & h_r^i & \cdots & h_r^L \\ \vdots & & \vdots & & \vdots \\ h_R^1 & \cdots & h_R^i & \cdots & h_R^L \end{bmatrix} \quad (3)$$

where h_r^i represents the RSS between i -th AP and r -th RP. When the RSS database construction is completed, the offline step ends. Afterward, the online step is a step in which the positioning target exists in the environment and the positioning is performed by applying a matching technique. In this paper, the double deep Q-Network method is proposed as a positioning method.

4.2. Cell-ID Method

In this subsection, we describe the Cell-ID method and propose its convergence method with fingerprinting. The conceptual diagram of the positioning system using the Cell-ID technique is shown in Figure 4.

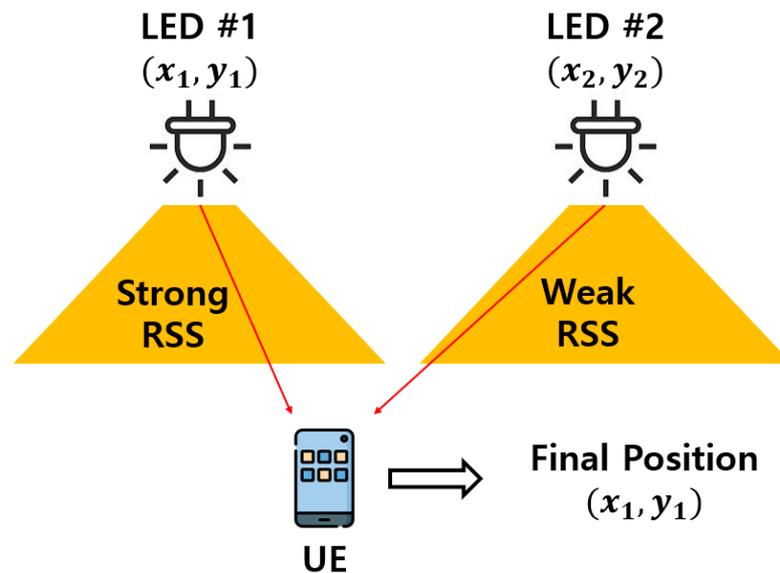


Figure 4. Conceptual diagram of Cell-ID method.

As shown in Figure 4, when there are two LED APs and one UE, the UE receives signals from both APs. In most cases, the signal received by the UE can be classified as either a strong signal or a weak signal. At this time, Cell-ID-based positioning determines the position of the UE based on the location of the LED AP that offers a stronger signal.

Hence, the Cell-ID method is one of the simplest positioning schemes. In the figure, # represents ID of AP.

However, this method only determines the approximate location of the UE, making it challenging to achieve high precision. Therefore, it can be considered to apply it in combination with other positioning methods. In this study, the initial state of the DDQN agent is determined using the positioning results of the Cell-ID technique. The double deep Q-Network is explained in detail in the following subsections.

4.3. Double Deep Q-Network

In this subsection, the use of the double deep Q-Network for positioning is explained. The double deep Q-Network model is developed to improve the issue of overestimation encountered in conventional DQN models [25]. Briefly, the problem of overestimations is that the agent always chooses the non-optimal action in any given state only because it has the maximum Q-value. In basic Q-learning, the optimal policy of the agent is always to choose the best action in any given state. The assumption behind the idea is that the best action has the maximum expected/estimated Q-value. However, the agent knows nothing about the environment in the beginning, so it needs to estimate $Q(s,a)$ at first and update it at each iteration. Such Q-values have lots of noise, and we are never sure whether the action with the maximum expected/estimated Q-value is really the best one. Unfortunately, the best action often has smaller Q-values than the non-optimal ones in most cases. According to the optimal policy in basic Q-learning, the agent tends to take the non-optimal action in any given state only because it has the maximum Q-value. Such a problem is called the overestimation of the action value (Q-value). When such a problem occurs, the noises from the estimated Q-value will cause large positive biases in the updating procedure. As a consequence, the learning process will be very complicated and messy. Hence, the authors proposed a solution to decouple the action selection during evaluation from the action taken by the agent. Consequently, instead of using the Bellman equation used in the DQN algorithm, (4) is used:

$$Q(s, a; \theta) = r + \gamma Q(s', \operatorname{argmax}_{a'} Q(s', a'; \theta); \theta') \quad (4)$$

As can be seen from (4), the main parameter of the neural network determines the optimal next action among all the available next actions. The target neural network can then evaluate that action to know the Q-value. This result can solve the problem of overestimation and provide a better final policy. Based on these advantages, this paper proposes indoor positioning by implementing the double deep Q-Network instead of the DQN.

Now, the factors considered in this study for reinforcement learning—the agent, state, action, and reward—will be explained. The agent predicts the location of the UE based on the RSS. The state represents the fingerprint data for the predicted point location. At this time, the initial state is determined based on the UE's RSS. Action refers to the agent moving the predicted point location up, down, left, or right. The reward represents the agent's reward for its actions. Next, state, action, and reward are explained in detail.

4.3.1. State

In this study, state means the location information of prediction points estimated by the agent. The location information means the RP number. Therefore, the size of the state is the same as the total number of RPs. The agent can know the RSS about the estimation point in the fingerprint database by using the RP number, which represents the current state. Through this, the reward can be calculated, and the location of the UE can be determined.

In addition, to minimize the agent's positioning time, the initial state is determined by the coordinates of the AP that sends the strongest signal to the UE. Through this, the number of steps that the agent moves the estimation point can be reduced.

4.3.2. Action

The agent performs an action to determine the location of the UE, which means movement of the estimation point in this method. In this study, a total of 5 actions were defined. The 5 actions are as follows:

- Up = $P_E(t + 1) = n_r + RP_x$;
- Down = $P_E(t + 1) = n_r - RP_x$;
- Left = $P_E(t + 1) = n_r - 1$;
- Right = $P_E(t + 1) = n_r + 1$;
- Stop = $P_E(t + 1) = P_E(t)$.

where P_E represents the estimated point location of the agent, n_r represents the current state, and RP_x represents the total number of RPs placed on the x-axis.

4.3.3. Reward

Compensation evaluates the action performed by the agent. If the agent moves the prediction point closer to the UE, the behavior is reinforced with a positive reward. Conversely, if the agent moves the prediction point further away, the behavior is weakened with a negative reward. Here, the Euclidean distance between the RSS value of the UE and the fingerprinting database was used to calculate the compensation. The Euclidean distance between the previous state and the current state was calculated. The model's behavior was reinforced with a positive reward when the Euclidean distance decreased and a negative reward when the Euclidean distance increased. The reward r can be calculated as (5):

$$r = \sqrt{(X_c - X_p)^2 - (Y_c - Y_p)^2} \tag{5}$$

where X_c and X_p represent the x-coordinates of the current and previous states, respectively, and Y_c and Y_p represent the y-coordinates of the current and previous states, respectively.

4.3.4. Network Architecture

The following subsection describes the architecture of the double deep Q-Network model. The double deep Q-Network model used in this study is shown in Figure 4.

As shown in Figure 5, the double deep Q-Network network has the same architecture as the DQN network. As mentioned earlier, by separating action selection and evaluation, we avoid estimating the value of the current state through direct max Q. This can be expressed as sending $\text{argmax}_{a'} Q(s', a'; \theta)$ from the Q-Network to the target network.

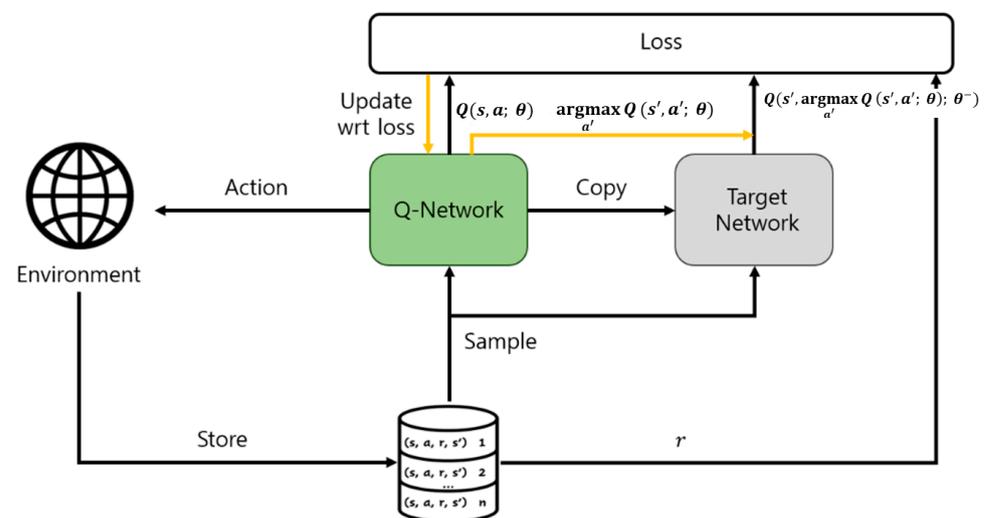


Figure 5. Architecture of DDQN model.

Additionally, the structure of the deep neural network utilizes a total of 3 dense layers and 1 flattened layer. The agent’s state serves as the input for the first layer, and the action is derived from the final layer. Two dense layers are employed as hidden layers in the neural network, with the rectifier linear unit (ReLU) applied as the activation function for each layer.

4.4. Block Diagram of Proposed Method

This section describes the block diagram of the proposed method in detail. The block diagram is shown in Figure 6.

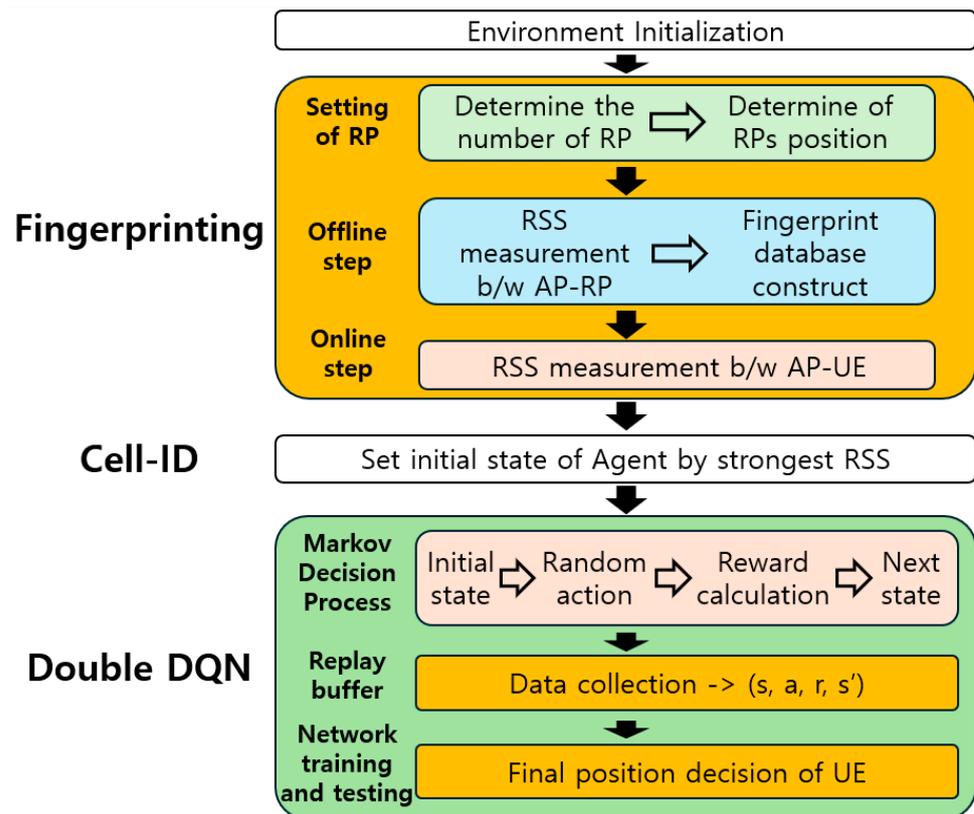


Figure 6. Block diagram of proposed method.

In Figure 6, the proposed method is divided into three steps. The first step is the fingerprinting. The core technology of the fingerprinting step is to determine the location of the RP and construct a fingerprinting database. The second step is to determine the initial state of the agent to reduce the processing time for positioning. For this purpose, the X and Y coordinates of the AP providing the strongest RSS were determined as the agent’s initial state. The third step is the double deep Q-Network algorithm, which is the core of this study. The agent performs the 5 actions defined in the above section and receives a reward accordingly. Based on the reward provided, it learns the policy to select the optimal action. The double deep Q-Network model that has been taught outputs the final location of the UE.

5. Simulation and Results

In this section, we describe the simulation parameters and present the results. The indoor VLC environment was implemented using MATLAB 2017b, while the double deep Q-Network model was designed based on Python 3.7 and implemented using the keras-rl2 library. The simulation environment and parameters are summarized in Table 2.

Table 2. Simulation parameters.

Parameter		Value
Indoor Environment	Room size	5 m × 5 m × 3 m
	No. APs	4
	No. RPs	676
	Episode	500
Double DQN	Max step of each episode	100
	Replay memory size	81,920
	Update of target network	100
	Policy	Epsilon Greedy

In Table 2, the RPs were positioned at 20 cm intervals within the indoor environment, with a total of 676 RPs set. For the double deep Q-Network model, a total of 500 episodes were run, with each episode limited to a maximum of 100 steps. Then, the replay memory size is set to 81,920. The weights of the target network were updated every 100 episodes in the Q-Network. The learning policy employed the Epsilon-Greedy policy. This policy is a strategy for making optimal decisions at each step without considering the future. In other words, it is an algorithm that assumes that the best choice at each step will lead to the best overall outcome. Of course, it does not always return the best results because it does not take future values into consideration.

Table 3 evaluates the learning performance by varying the number of nodes per layer in the designed reinforcement learning model. In this study, the total number of neural network layers was set to three, and a rectified linear unit (ReLU) was applied as the activation function for each layer. Indicators for assessing learning performance include average test scores and learning time. As shown in Table 3, as the number of nodes per layer increases, the learning time also increases, and the average test score improves. However, the learning time is 30 to 60 s longer than other combinations. This will require more learning time in the future if the total number of episodes increases or the search area expands. Conversely, when the number of nodes per layer is small, the average test score is low, but the learning time is the fastest.

Table 3. Performance comparison based on layer combinations.

	Parameter			Average Test Score	Learning Time
	1-Layer	2-Layer	3-Layer		
Layer	16	16	16	4.7	258.9 s
	32	32	32	8.0	259.8 s
	64	64	64	8.78	262.9 s
	128	128	128	9.78	317.9 s
	128	64	32	7.12	294.3 s
	64	32	16	9.0	282.5 s
	32	64	32	7.32	288.3 s

Therefore, in this study, the number of nodes per layer was selected as 64-32-16 based on the results in Table 3. As shown in the table, with 64-32-16 nodes per layer, a relatively high average test score of 9.0 can be achieved within a learning time of 282.5 s. Figure 7 shows the performance graph of learning through the model.

In Figure 7, it can be observed that the reward value rapidly converges to a value greater than 0 as the episode progresses in the double deep Q-Network model. This signifies that the agent is able to locate the adjacent RP to the UE more efficiently as the training proceeds. To test the positioning performance, double deep Q-Network agent training was performed for a total of 500 episodes. Utilizing the trained double deep Q-Network agent, as described earlier, a positioning test was conducted. For this purpose, the trained model was saved in the h5 format of the Keras library, a Python package. To test the saved model, the UE was deployed in approximately 1000 new episodes. Afterward, the positioning

errors and processing times that occurred during 1000 episodes were averaged. As a result, it was verified that the nearest RP to the UE was identified in approximately 0.03 s. The positioning error achieved was less than 13 cm, demonstrating good performance.

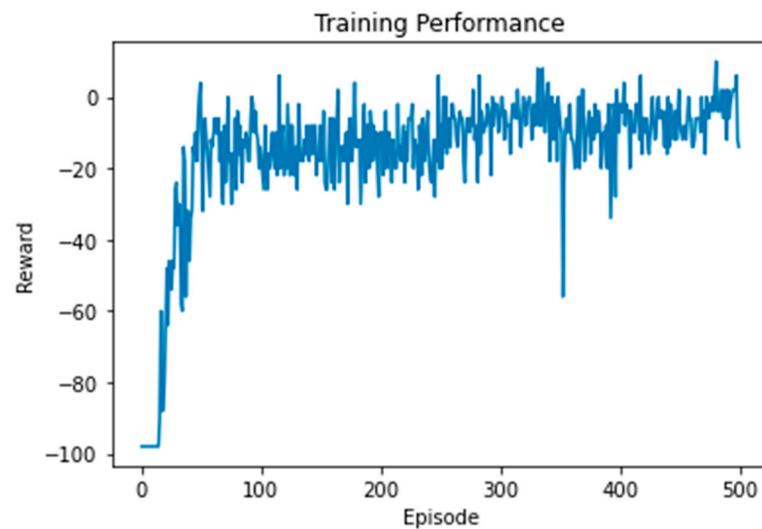


Figure 7. Training performance of double deep Q-Network model.

In Table 4, the performance of the proposed method is compared with existing methods. At this time, since the proposed method and KNN depend on the number of reference points, the same number of reference points was used. A total of 676 reference points were used, as shown in Table 2. Triangulation was determined using three APs that provide a strong RSS.

Table 4. Comparison between proposed and conventional scheme.

Performance	FP-Based DDQN (Proposed)	KNN	Triangulation
Positioning Error [m]	0.13	0.191	0.3281
Processing Time [s]	0.03	0.142	1.298

As shown in the table, all three techniques achieve accuracy within 1 m. Among them, Triangulation has the lowest positioning precision at 0.3281 m, while the proposed method shows the highest precision at 0.13 m. At this time, it was confirmed that a fast processing time of 0.03 s was achieved. This shows that the proposed scheme rapidly converges to the optimal value by limiting the initial state of the agent based on Cell-ID. Hence, based on these results, the effectiveness of the proposed method was verified.

6. Conclusions

Recently, due to the increasing complexity and size of indoor environments, the LBS industry has experienced rapid growth. Achieving precise positioning is crucial for delivering high-quality LBSs. Hence, this paper focuses on applying an AI algorithm based on a double deep Q-Network model for user positioning in VLC-based indoor environments. The proposed method combines the existing fingerprinting technique with the double deep Q-Network model. The main components of this study include analyzing the channel characteristics of the indoor VLC environment, constructing a fingerprinting map, and applying the double deep Q-Network model. This paper emphasizes the optimal matching based on the fast processing time of the double deep Q-Network model. For this purpose, each element of reinforcement learning is defined and mathematical modeling

is performed. The double deep Q-Network model is trained through simulations for approximately 100 episodes. By utilizing the trained positioning model, it is shown that it achieves a positioning accuracy of 13 cm and just 0.03 s of time duration.

In the future, there are plans to explore additional deep reinforcement learning models, such as dueling DQN and the deep deterministic policy gradient (DDPG). Although the processing time in the online step of the fingerprinting was reduced in this study, the issue of the cost of constructing the database in the offline step still remains. Furthermore, it is also considered to enhance the learning process by incorporating semi-supervised reinforcement learning, which utilizes a small amount of data for training.

Author Contributions: Conceptualization, S.H.O. and J.G.K.; methodology, S.H.O. and J.G.K.; software (MATLAB 2017b, Python 3.7, Keras-rl2), S.H.O.; validation, S.H.O. and J.G.K.; formal analysis, S.H.O. and J.G.K.; investigation, S.H.O.; resources, J.G.K.; data curation, S.H.O.; writing—original draft preparation, S.H.O.; writing—review and editing, J.G.K.; visualization, S.H.O.; supervision, J.G.K.; project administration, J.G.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2021R1F1A1063845).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hou, Z.; Kim, K.; Zhang, G.; Li, P. A Study on the Realization of Virtual Simulation Face Based on Artificial Intelligence. *J. Inf. Commun. Converg. Eng.* **2023**, *21*, 152–158. [[CrossRef](#)]
2. Lee, S.; Park, H.; You, Y.; Yong, S.; Moon, I.K. Effects of CNN Backbone on Trajectory Prediction Models for Autonomous Vehicle. *J. Inf. Commun. Converg. Eng.* **2023**, *21*, 346–350. [[CrossRef](#)]
3. Zafari, F.; Gkelias, A.; Leung, K.K. A Survey of Indoor Localization Systems and Technologies. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2568–2599. [[CrossRef](#)]
4. Hsu, L.T.; Gu, Y.; Kamijo, S. 3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation. *GPS Solut.* **2016**, *20*, 413–428. [[CrossRef](#)]
5. Li, N.; Chen, J.; Yuan, Y. A WiFi indoor localization strategy using particle swarm optimization based artificial neural networks. *Int. J. Distrib. Sens. Netw.* **2016**, *2016*, 4583147. [[CrossRef](#)]
6. Shang, S.; Wang, L. Overview of WiFi fingerprinting-based indoor positioning. *IET Commun.* **2022**, *16*, 725–733. [[CrossRef](#)]
7. Alarifi, A.; Al-Salman, A.; Alsaleh, M.; Alnafessah, A.; Al-Hadhrami, S.; Al-Ammar, M.A.; Al-Khalifa, H.S. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors* **2016**, *16*, 707. [[CrossRef](#)] [[PubMed](#)]
8. Wang, Q.; Feng, Y.; Zhang, X.; Sun, Y.; Lu, X. IWKNN: An effective Bluetooth positioning method based on Isomap and WKNN. *Mob. Inf. Syst.* **2016**, *2016*, 8765874. [[CrossRef](#)]
9. Luo, J.; Fan, L.; Li, H. Indoor positioning systems based on visible light communication: State of the art. *IEEE Comm. Surv. Tutor.* **2017**, *19*, 2871–2893. [[CrossRef](#)]
10. Seguel, F.; Krommenacker, N.; Charpentier, P.; Soto, I. Visible light positioning based on architecture information: Method and performance. *IET Commun.* **2019**, *13*, 848–856. [[CrossRef](#)]
11. Lian, J.; Vatansever, Z.; Noshad, M.; Brandt-Pearce, M. Indoor visible light communications, networking, and applications. *J. Phys. Photonics* **2019**, *2019*, 012001. [[CrossRef](#)]
12. Priya, C.B.; Sivakumar, S. A survey on localization techniques in wireless sensor networks. *Int. J. Eng. Technol.* **2017**, *7*, 125–129. [[CrossRef](#)]
13. Zhao, C.; Zhang, H.; Song, J. Fingerprint and Visible Light Communication Based Indoor Positioning Method. In Proceedings of the 9th International Conference on Advanced Infocomm Technology (ICAIT), Chengdu, China, 22–24 November 2017.
14. Dou, F.; Lu, J.; Xu, T.; Huang, C.-H.; Bi, J. A Bisection Reinforcement Learning Approach to 3-D Indoor Localization. *IEEE Internet Things J.* **2021**, *8*, 6519–6535. [[CrossRef](#)]
15. Yang, T.; Cabani, A.; Chafouk, H. A Survey of Recent Indoor Localization Scenarios and Methodologies. *Sensors* **2021**, *21*, 8086. [[CrossRef](#)] [[PubMed](#)]
16. Yoo, J. Multiple Fingerprinting Localization by an Artificial Neural Network. *Sensors* **2022**, *22*, 7505. [[CrossRef](#)]

17. Sayed, W.; Ismail, T.; Elsayed, K. A Neural Network-Based VLC Indoor Positioning System for Moving Users. In Proceedings of the International Conference on Smart Applications, Communications and Networking (SmartNets), Sharm El Sheikh, Egypt, 17–19 December 2019.
18. Van, M.T.; Tuan, N.V.; Son, T.T.; Le-Minh, H.; Burton, A. Weighted k-nearest neighbor model for indoor VLC positioning. *IET Commun.* **2017**, *11*, 864–871. [[CrossRef](#)]
19. Tran, H.Q.; Ha, C. Fingerprint-Based Indoor Positioning System Using Visible Light Communication—A Novel Method for Multipath Reflections. *Electronics* **2019**, *8*, 63. [[CrossRef](#)]
20. Xu, S.; Chen, C.-C.; Wu, Y.; Wang, X.; Wei, F. Adaptive Residual Weighted K-Nearest Neighbor Fingerprint Positioning Algorithm Based on Visible Light Communication. *Sensors* **2020**, *20*, 4432. [[CrossRef](#)] [[PubMed](#)]
21. Zhang, Z.; Chen, H.; Zeng, W.; Cao, X.; Hong, X.; Chen, J. Demonstration of Three-Dimensional Indoor Visible Light Positioning with Multiple Photodiodes and Reinforcement Learning. *Sensors* **2020**, *20*, 6470. [[CrossRef](#)]
22. Zhang, Z.; Zhu, Y.; Zhu, W.; Chen, H.; Hong, X.; Chen, J. Iterative point-wise reinforcement learning for highly accurate indoor visible light positioning. *Opt. Express* **2019**, *27*, 22161–22172. [[CrossRef](#)]
23. Xu, M.; Xia, W.; Jia, Z.; Zhu, Y.; Shen, L. A VLC-Based 3-D Indoor Positioning System Using Fingerprinting and K-Nearest Neighbor. In Proceedings of the IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, Australia, 4–7 June 2017.
24. Martínez-Ciro, R.A.; López-Giraldo, F.E.; Luna-Rivera, J.M.; Ramírez-Aguilera, A.M. An Indoor Visible Light Positioning System for Multi-Cell Networks. *Photonics* **2022**, *9*, 146. [[CrossRef](#)]
25. Hasselt, H.V.; Guez, A.; Silver, D. Deep Reinforcement Learning with Double Q-Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.

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