

Article

Adaptive Volt–Var Control in Smart PV Inverter for Mitigating Voltage Unbalance at PCC Using Multiagent Deep Reinforcement Learning

Yoongun Jung ¹, Changhee Han ¹ , Dongwon Lee ¹ , Sungyoon Song ² and Gilsoo Jang ^{1,*} 

¹ School of Electrical Engineering, Korea University, Anam-ro, Sungbuk-gu, Seoul 02841, Korea; mundd13@korea.ac.kr (Y.J.); hch0806@korea.ac.kr (C.H.); dong951106@korea.ac.kr (D.L.)

² Advanced Power Grid Research Center, Korea Electrotechnology Research Institute (KERI), 138, Naesonsunhwan-ro, Uiwang-si 16029, Korea; songsy@keri.re.kr

* Correspondence: gjang@korea.ac.kr; Tel.: +82-3290-3246; Fax: +82-2-3290-3692

Abstract: Modern distribution networks face an increasing number of challenges in maintaining balanced grid voltages because of the rapid increase in single-phase distributed generators. Because of the proliferation of inverter-based resources, such as photovoltaic (PV) resources, in distribution networks, a novel method is proposed for mitigating voltage unbalance at the point of common coupling by tuning the volt–var curve of each PV inverter through a day-ahead deep reinforcement learning training platform with forecast data in a digital twin grid. The proposed strategy uses proximal policy optimization, which can effectively search for a global optimal solution. Deep reinforcement learning has a major advantage in that the calculation time required to derive an optimal action in the smart inverter can be significantly reduced. In the proposed framework, multiple agents with multiple inverters require information on the load consumption and active power output of each PV inverter. The results demonstrate the effectiveness of the proposed control strategy on the modified IEEE 13 standard bus systems with time-varying load and PV profiles. A comparison of the effect on voltage unbalance mitigation shows that the proposed inverter can address voltage unbalance issues more efficiently than a fixed droop inverter.

Keywords: voltage unbalance; volt–var curve control; smart PV inverter; multiagent proximal policy optimization



Citation: Jung, Y.; Han, C.; Lee, D.; Song, S.; Jang, G. Adaptive Volt–Var Control in Smart PV Inverter for Mitigating Voltage Unbalance at PCC Using Multiagent Deep Reinforcement Learning. *Appl. Sci.* **2021**, *11*, 8979. <https://doi.org/10.3390/app11198979>

Academic Editor: Alfio Dario Grasso

Received: 20 August 2021

Accepted: 21 September 2021

Published: 26 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

The effort in decarbonizing a power system is undergoing major changes including a large amount of distributed generation (DG) replacing conventional generators and transportation undergoing electrification. Most DG, such as photovoltaic (PV) generation and electric vehicle (EV) charging stations, are connected to a distribution network in a single phase. The high penetration of DG and EV charging stations has resulted in new challenges for maintaining the proper voltage quality in a distribution network. The fluctuating characteristic output patterns of single-phase DG and the irregularity of EV charging patterns [1,2] aggravated the voltage unbalance. Unbalanced voltages cause heat, vibration, and inefficiency by shortening the lifespan of three-phase transformers and power-electronics-based equipment [3,4]. Especially at the point of common coupling (PCC) in the distribution network (DN), it is necessary to reduce the voltage unbalance because it is directly connected to the high-voltage stage with a three-phase transformer.

Various methods have been proposed to mitigate the voltage unbalance at high PV penetration rates in distribution networks. Power-electronics-based equipment, such as static synchronous compensators or passive devices, such as shunt capacitors, can mitigate the unbalanced voltage by compensating for the reactive power. However, these methods have low economic feasibility because they require additional equipment and maintenance

costs. Most strategies proposed for using PV inverters to improve phase unbalance rely on centralized cooperative control to solve the economic problem. For example, [5] presents a bi-level VVO framework for CVR by coordinating the operation of distribution system's legacy voltage control devices and smart inverters to efficiently handle the discrete and continuous control variables. Ref. [6] analyses and compares four of the most relevant smart EV charging controls as follows: (1) active power droop control; (2) reactive power droop control for single-phase EV chargers; (3) Load Balancing control; and (4) Sequence Compensation control (SC), with the aim of reducing the VUF and avoiding under-voltage conditions. In a previous article [7], a method for mitigating voltage unbalance by adjusting reactive power injections from PV systems was presented. By using the Karush–Kuhn–Tucker optimization method, the negative and zero phase components of the voltage can be selectively reduced. In another study [8], the voltage unbalance in the DN was solved using the reactive power compensation capability of the inverter through centralized and decentralized methods. The centralized method solves the voltage unbalance by presenting a mathematically calculated value using a Steinmetz design. However, the process of centralized control requires three steps: (1) receiving measured data through communication; (2) calculating set points based on their optimization problem, which is formulated to obtain some optimal parameters with the predefined objective functions; and (3) sending the set points to each control unit. These procedures are repeated every control time. In the process, the load on the system or the output of renewable energy has already changed. And computation time is accompanied by a correlation between the scale of the power system and the computational load. Thus, centralized controls are difficult to dispatch accurately in real time.

On the other hand, the Deep Reinforcement Learning (DRL) method calculates the optimal parameters based on offline analysis results and uses them for the next day's online operation. Moreover, the optimal policies which are composed by neural network are determined by performing power flow calculations on the previous day. The Neural networks (NNs) take little time (less than 1ms) to derive the output, so NN has an advantage in using real time [9]. This feature can be usefully applied when the fast response in the parameter updates of the control unit is required for online operation. Decentralized control adopting DRL enables many system operating points to respond in real time with little communication. Therefore, recent efforts have focused on problems in power systems combined with DRL [10–15]. In several challenging tasks for power system operation, DRL shows great potential. For example, a novel adaptive emergency control scheme using DRL by leveraging the high-dimensional feature extraction and nonlinear generalization capabilities of DRL for complex power systems was proposed [14]. Elsewhere [15], the volt-var optimization algorithm was proposed for reinforcing voltage regulation and power loss reduction. However, the existing voltage control methods using DRL only focus on adjusting voltage profiles without looking into mitigating voltage unbalance and are not flexible enough to adopt changing distribution networks with various newly connected devices and inverter-based resources.

In this article, a novel volt-var control strategy for single-phase DG smart inverters for mitigating voltage unbalance targeting the PCC under time-varying operating conditions in a three-phase distribution system is proposed. The problem is extended to the Markov decision process by integrating it with the DRL algorithm. Multiagent proximal policy optimization methods [16] were used to train multiple inverters in a distribution network owing to the stable and fast convergence ability. In the proposed framework, a probability model based on actual PV generations and load profiles is adopted to ensure that the inverter acts properly under the influence of uncertainty in PV outputs. Then, a voltage-passive device, such as the on-load tap-changer (OLTC) and static voltage compensator, is considered as a realistic assumption for a grid. Furthermore, because more inverter-based resources will be added to the grid in the future, an adaptive learning method is needed by using the latest learning models, which are trained using the digital twin model [17]. The digital twin provides an environment for exploration and exploitation for the learning of

agents that converges to the optimal solution through various experiences. The proposed volt-var control, which is used by the distribution system operator in control rooms, was predefined through day-ahead training based on forecast PV and load consumption data. Figure 1 shows the structure of updating the model through day-ahead training in data learning centers using digital twins.

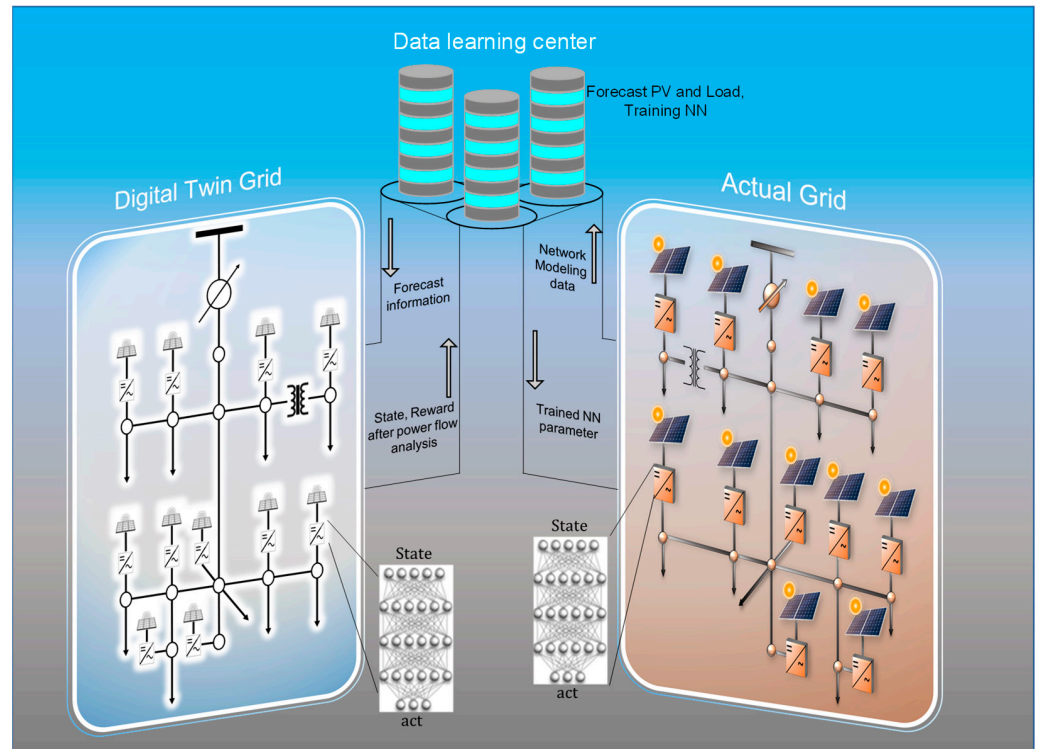


Figure 1. Day-ahead training in data learning center using a digital twin grid and implemented in an actual grid platform.

The procedure for using the developed platform mainly includes two stages: (1) at the training stage, day-ahead training of smart inverters with forecast information about PV generation and load consumption; and (2) at the implementation stage, for using the NN trained on the actual grid, updating the existing NN to the trained one. During the training stage, agents learn an optimal policy through exploration and exploitation and automatically save the best-performance NN parameters. Once the training stage is completed, the agents at the implementation stage use the trained optimal policy to provide optimal control actions to the environment.

These efforts are detailed in the following sections. Section 2 shows the voltage unbalance factor at PCC, the volt-var curve and the reactive power capability of each inverter. Section 3 shows the Multiagent DRL Design for Proposed Method, and Section 4 shows the simulation result carried out in the modified IEEE 13-bus system with time-varying load and PV profiles. The conclusion and future work are presented in Section 5.

2. Problem Formulation

2.1. Voltage Unbalance Factors and Sequence Voltage

Distribution networks are experiencing a steady increase in the number of unbalanced components, such as single-phase PV systems, EV charging stations, and loads distributed along feeders that cause voltage unbalance at the PCC. The voltage unbalance is defined by voltage unbalance factors (VUFs) [18,19]. To calculate VUFs, sequence voltages such as positive sequence voltage V^+ , negative sequence voltage V^- , and zero-sequence voltage V^0 , are required, which are computed by the Fortescue transformation. Equation (1)

shows the Fortescue transformation, where $\alpha = e^{j2\pi/3} = 1\angle 120^\circ$, and V_A, V_B, V_C are phase-to-neutral voltages [20].

$$\begin{bmatrix} V^0 \\ V^+ \\ V^- \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \alpha^2 & \alpha \\ 1 & \alpha & \alpha^2 \end{bmatrix}^{-1} \begin{bmatrix} V_A \\ V_B \\ V_C \end{bmatrix} \tag{1}$$

The negative sequence voltage unbalance factor and the zero-sequence voltage unbalance factor are defined as VUF^- and VUF^0 , where $VUF^-(\%) = \frac{|V^-|}{|V^+|} \times 100$ and $VUF^0(\%) = \frac{|V^0|}{|V^+|} \times 100$. In this study, the aim was to minimize the total sum of VUF^- and VUF^0 at the PCC during one day. The objective function defined by the VUFs is explained in Section 3.3.

2.2. Volt–Var Curve

In the proposed strategy, mitigating the voltage unbalance at the PCC is achieved by passively injecting or absorbing reactive power into the points connected in the PV system by following the defined volt–var curve, as illustrated in Figure 2, without altering the active power injection, where the x -axis represents the phase to a neutral voltage at the bus, which is connected to the PV system. The IEEE 1547 standard provides a minimum requirement guideline for the volt–var curve, where $V_1 = 0.92 \times V_{ref}$, $V_2 = 0.98 \times V_{ref}$, $V_3 = 1.02 \times V_{ref}$, and $V_4 = 1.08 \times V_{ref}$ for DER in category B [21]. The proposed scheme adjusts the scales of V_1, V_2, V_3, V_4 . When agents observe the state, the output of the NN is a combination of V_1, V_2, V_3, V_4 . After training, each NN creates an optimal volt–var curve in a given situation. $Q_{i,t}^{lim}$ is the reactive power capability at time t for i -th PV system.

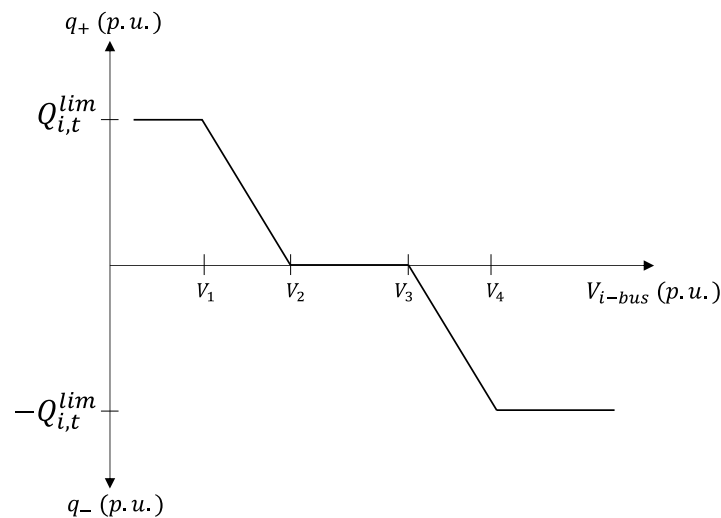


Figure 2. Volt–var curve for each inverter.

2.3. Reactive Power Capability

The reactive power capability for each PV system is determined by its apparent power rating S_i^{rated} and the active power rating P_i^{rated} for the i -th PV system, as shown in Equation (2), where $kvarMax$ is the maximum value of the reactive power per unit that the inverter can provide to the grid. If the i -th PV inverter is less than $0.05 \times P_i^{rated}$, it is not allowed to provide/absorb reactive power. In addition, it is not allowed to provide/absorb reactive power more than $kvarMax$. According to the IEEE 1547 standard, it is assumed that the magnitude of $kvarMax$ is $0.44 \times S_i^{rated}$. Figure 3 shows the reactive power capability curve.

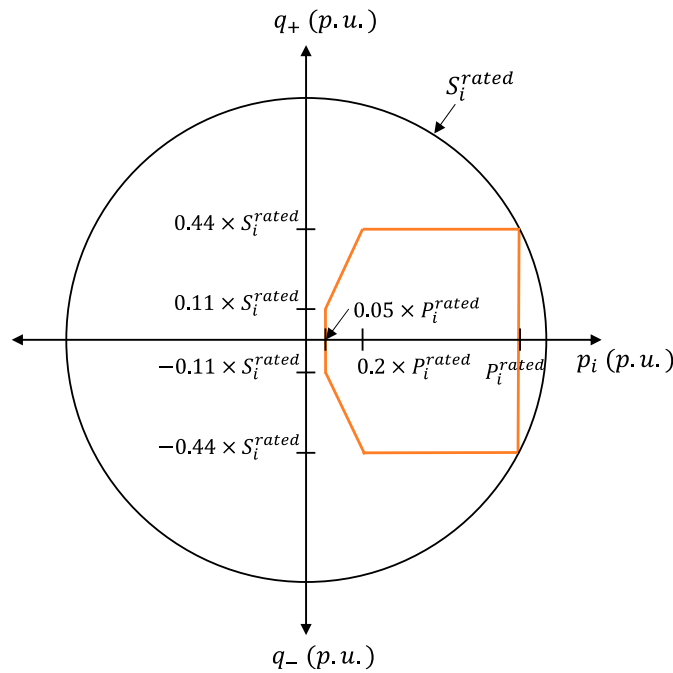


Figure 3. Reactive power capability curve for each inverter.

$$Q_{i,t}^{lim} = \begin{cases} 0, & 0 \leq P_{i,t} \leq 0.05 \times P_i^{rated} \\ \frac{P_{i,t}}{P_i^{rated}} \times kvarMax, & 0.05 \times P_i^{rated} < P_{i,t} \leq 0.2 \times P_i^{rated} \\ kvarMax, & 0.2 \times P_i^{rated} < P_{i,t} \end{cases} \quad (2)$$

3. Multiagent DRL Design for Proposed Method

3.1. Principles of Deep Reinforcement Learning

In reinforcement learning (RL), the multiple agents learn to make optimal decisions by interacting with the environment through exploration and exploitation, as illustrated in Figure 4, with a set of states (S_t), actions (A_t) and rewards (R_t) for each agent. At time t , each agent observes the state and receives the reward from the environment. At the same time, each agent takes action based on its policy.

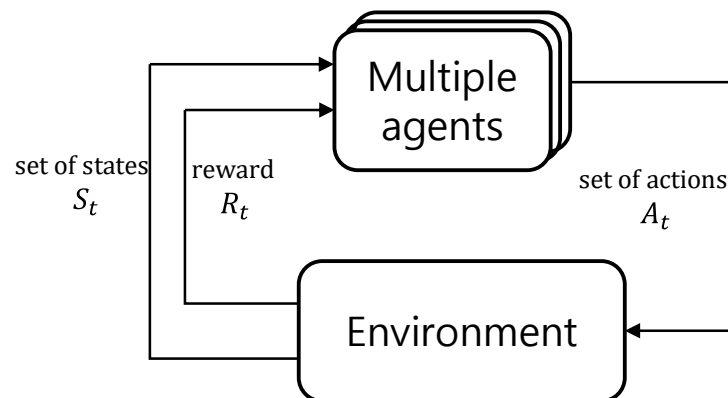


Figure 4. Framework of multiagent RL.

The goal is to find an optimal action for each agent in the current state based on their policy, where $\gamma^t \in [0, 1]$ is a discount factor. Thus, solving RL involves finding the individual policy expressed as $\pi_\theta(a|s)$, which is parameterized by θ , which is determined

by the weights and biases of an NN. The agent updates θ to maximize the cumulative discounted rewards G_t , as shown in Equation (3).

$$G_t = \sum_{t=0}^T \gamma^t r_t \quad (3)$$

The state value function $V^\pi(s)$ and action-value function $Q^\pi(s, a)$ are two important value functions in RL, which are defined in Equations (4) and (5), respectively. Here, $V^\pi(s)$ calculates the expected reward for agents starting in such a state according to a particular policy to evaluate the value of a particular state. However, $Q^\pi(s, a)$ calculates the expected reward for agents starting in such a state with action according to a particular policy to evaluate the value of an action.

$$V^\pi(s) = E(G_t | s_t = s; \pi) \quad (4)$$

$$Q^\pi(s, a) = E(G_t | s_t = s, a_t = a; \pi) \quad (5)$$

3.2. Proximal Policy Optimization

Proximal policy optimization (PPO) is based on the Monte Carlo policy gradient method, which computes an estimator of the policy gradient (PG) using a stochastic gradient ascent algorithm. An estimated return form of the PG algorithm was reported in a previous article to update the policy parameter θ iteratively using the Monte Carlo method. Because the total expected return G_t can be calculated from the real sample trajectory, the policy gradient is updated as

$$\nabla_{\theta} J(\theta) = \nabla_{\theta} V^{\pi}(s_0)(\theta) = E_{\pi}[Q^{\pi}(s, a) \nabla_{\theta} \ln \pi_{\theta}(a|s)] \quad (6)$$

$$\theta \leftarrow \theta + \alpha \nabla_{\theta} J(\theta) \quad (7)$$

where α is the learning rate, which can be tuned manually, and $J(\theta)$ is the objective function.

However, the PG algorithm does not guarantee adequate convergence, so PPO secures it in two ways: (1) using the advantage function of Equation (8) instead of the action-value function $Q^\pi(s, a)$ of Equation (6) to reduce the variance of the estimate during parameter update and (2) adopting a certain trust region to improve the stability of training by ensuring a clipped surrogate objective at every iteration by modifying $J(\theta)$ in Equation (6) to Equations (9) and (10).

$$Ad^{\pi}(s, a) = Q^{\pi}(s, a) - V^{\pi}(s) \quad (8)$$

$$J(\theta) = E_{\pi_{\theta_{old}}} \left[\min \left(\frac{\pi_{\theta}(a|s)}{\pi_{\theta_{old}}(a|s)} Ad^{\pi_{\theta_{old}}}(s, a), g(\epsilon, Ad^{\pi_{\theta_{old}}}(s, a)) \right) \right] \quad (9)$$

$$Jg(\epsilon, Ad^{\pi_{\theta_{old}}}(s, a)) = \begin{cases} (1 + \epsilon) Ad^{\pi_{\theta_{old}}}(s, a), & Ad^{\pi_{\theta_{old}}}(s, a) \geq 0 \\ (1 - \epsilon) Ad^{\pi_{\theta_{old}}}(s, a), & Ad^{\pi_{\theta_{old}}}(s, a) < 0 \end{cases} \quad (10)$$

where θ_{old} is the present parameters in the NN, and the size of the trust region for updates is decided by ϵ . By clipping operators, the new policy does not influence excessively to avoid being trapped in the local minimum of the objective function geometry. The convergence performance of the PPO algorithm becomes stable and effective when learning multiple agents through these two methods. For the reasons mentioned above, PPO was applied to the method of learning the PV inverter to mitigate the voltage unbalance at the PCC.

3.3. PPO-Based Multiagent DRL Framework for Autonomous Control

In the proposed training process, the environment is the DN, which is composed of time-varying load consumption and PV generation and passive OLTC. Massive episodes are applied to train inverters connected to PV systems. To address such errors as the forecast error of load consumption and PV production or errors occurring because of communication, each episode has Gaussian noise. Each agent has its own actor-critic NN,

which is trained to take actions with respect to the given operating condition to mitigate voltage unbalance at the PCC. Figure 5 shows the training architecture of the proposed learning platform. The states, actions, and rewards are discussed in this section.

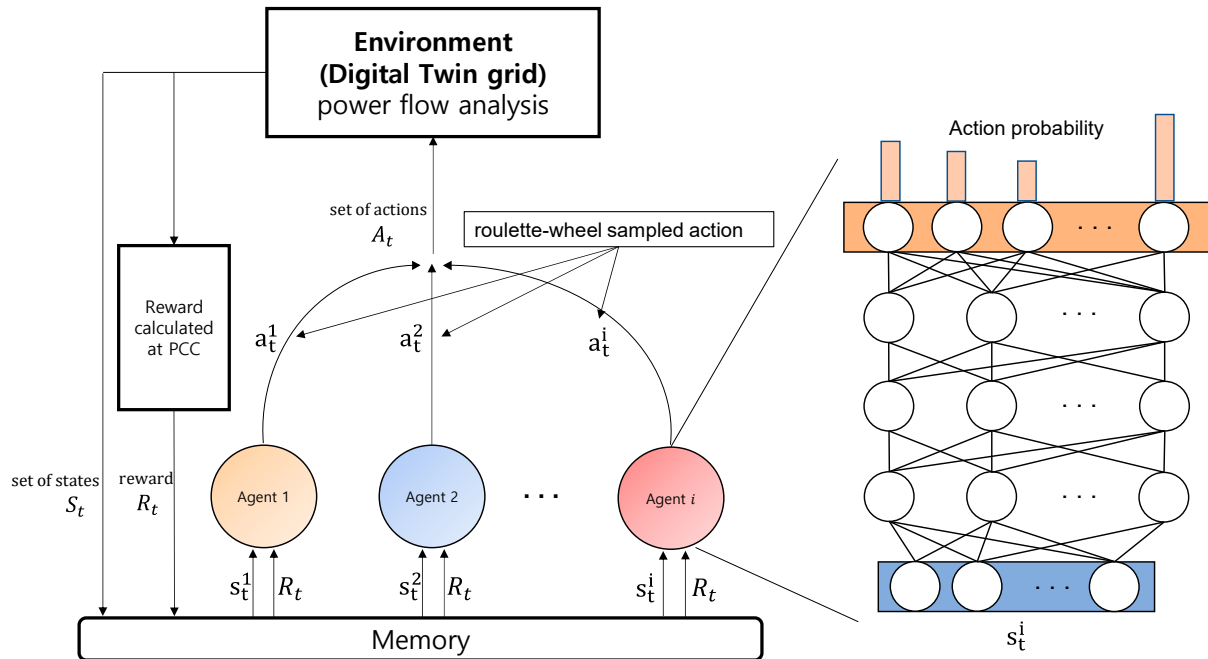


Figure 5. Training architecture for the proposed learning platform.

3.3.1. State

The appropriate choice of the state is important because each agent determines the action using the state. In this study, the PV and load have a large influence when determining the volt–var curve of each inverter, so the amount of PV generation and load is set as the state. In the case of day-ahead training for real-time operation, learning is carried out by putting the i -th PV generation and load as the state predicted on the previous day. The set of states for the i -th inverter at time step t is denoted as $s_t^i = \{P_t^{pv,i}, S_t^{load}\}$, where $P_t^{pv,i}$ represents the active power of the i -th PV generation, and S_t^{load} is the normalized apparent power of the load consumption. In addition, various pieces of system information can be helpful when determining the volt–var curve for the inverter; however, additional information is not used because the communication burden must be reduced for real-time control.

3.3.2. Action

To mitigate voltage unbalance at the PCC at each time step t , each inverter determines the volt–var curve according to its action. The action was the set of $\{V_1, V_2, V_3, V_4\}$, where $V_1 < V_2 < 1 < V_3 < V_4$. V_1, V_2 are selected in $\{0.92, 0.94, 0.96, 0.98\}$, and V_3, V_4 are selected in $\{1.02, 1.04, 1.06, 1.08\}$. Therefore, the number of action spaces is 36. When states pass through the NN of the agent, the probability of action is returned. The action for i -th inverter, represented as a_t^i , was selected by the roulette-wheel selection method [22] according to the probability. After the training, the NN assigns a large part of the probability to the optimal action.

3.3.3. Reward

A proper reward function is necessary to train the policy of the multiple agents appropriately. In this study, the reward function is defined as

$$R_t = -\{VUF^{proposed} - VUF^{fixed}\} \quad (11)$$

where $VUF^{proposed} = VUF_2^{proposed} + VUF_0^{proposed}$ and $VUF^{fixed} = VUF_2^{fixed} + VUF_0^{fixed}$ respectively. Because the policy of agents is trained to maximize the reward, the minus was implemented at the beginning of Equation (11) to minimize VUF at the PCC. In addition, $-VUF^{fixed}$ makes the proposed strategy more effective than fixed volt-var curve inverters. The fixed volt-var curve inverters are explained in Section 4.2.

3.4. Proposed Algorithm

The proposed training and the implementation algorithm are shown in Figure 6. Algorithm 1 presents the pseudocode of the training algorithm in detail. The closed-loop training dataset comprises 3000 episodes, where each episode comprises 50,400 1-s intervals (from 6 to 20 o'clock assuming that the PV system generates active power during daytime) of raw data. For each episode, the corresponding decisions and rewards are saved in batches to calculate the gradient descent to update the NN. After day-ahead training, NN has an optimal policy to act properly.

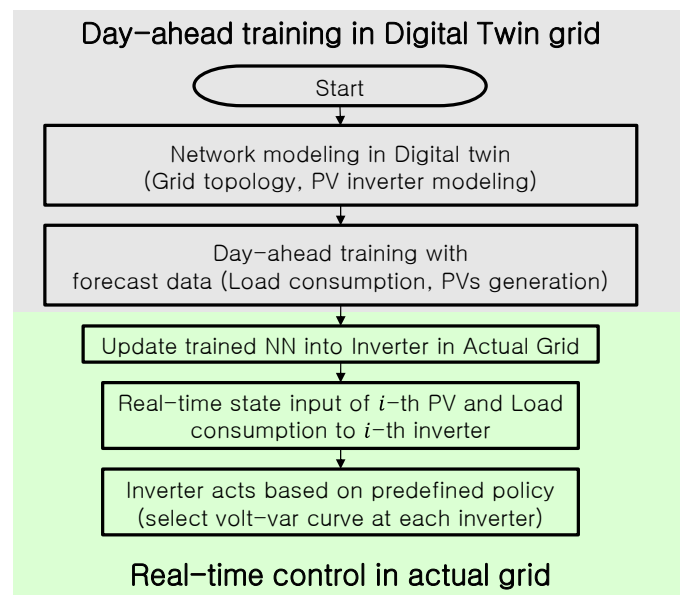


Figure 6. Information flow of the day-ahead training in the digital twin and implementation in the actual grid.

The PV generation data and load data were selected from the historical yearly PV generation data of the Korea Power Exchange [23]. Figure 7 shows the normalized PV generation and load consumption, which are applied to various PV generation and load consumption per unit.

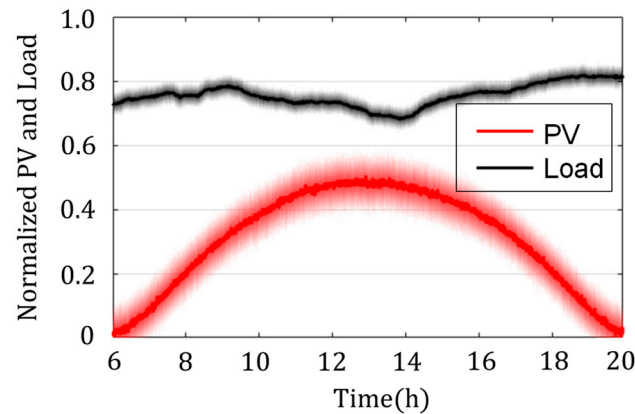


Figure 7. Normalized PV generation and load consumption profiles.

At each time step t (1 s), PV generation and load consumption, including their Gaussian noise, are used to consider the difference between the forecast and actual active power of the PV generation load consumption. The stochastic variations in PV generation and load consumption follow normal distributions. Equation (12) shows a normal distribution function.

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2} \quad (12)$$

where μ denotes the mean value of PV generation and load consumption at time t , and δ denotes the standard deviation of PV generation and load consumption.

Algorithm 1. Training algorithm based on PPO.

```

1: Initialize network
2: Initialize the critic and actor networks with weights
3: for episode = 1 to N do
4:   for time step = 360 min (6 h) to 1200 min (20 h) do
5:     each agent selects an action  $a_i^t$  for each  $s^t \in S$ 
6:     Execute  $a_i^t$  in environment
7:     for time step = 0 s (0 min) to 60 s (1 min) do
8:       Solve power flow with proposed volt-var curve
9:       Calculate  $VUF^{proposed}$  at PCC
10:      Solve power flow with fixed volt-var curve
11:      Calculate  $VUF^{fixed}$  at PCC
12:      reward calculate  $-\{VUF^{proposed} - VUF^{fixed}\}$ 
13:      end
14:      Send a set of states  $s^{t+1}$  and reward  $r^t$  to each agent
15:      Each agent selects  $a_i^{t+1}$  through  $s^{t+1}$ 
16:      Store the transition pairs in replay buffer
17:      Sample a random minibatch
18:      Update target critic and actor by stochastic gradient
19:      step+=1
20:    end
21:  end

```

4. Case Study

In the process of analyzing the voltage profiles in a distribution system, it is important to use an accurate continuous power flow calculation methodology. Recently, novel methodologies have been proposed for solving both Power-Flow and Continuation Power-Flow in distribution systems. Reference [24] proposed several corrector techniques based on efficient Newton-like methods to improve the computational performance of conventional continuation power flow analysis. Reference [25] studied some relevant

aspects related with the power flow solution of ill-conditioned cases using the current injection formulation. Reference [26] solve the OPF problem using meta-heuristic MPA algorithm. In this study, we used the OpenDSS as the power flow calculation tool which is an electric power distribution system simulator developed by the Electric Power Research Institute [27]. OpenDSS has great advantage that determines the voltage and operating point of each inverter based on its volt-var curve. Moreover, OpenDSS offers a Python COM interface that enables users to script customized simulations adopting Python’s reinforcement learning package PyTorch, which provides NN and training for agents [28]. To demonstrate the effectiveness of the system, a simulation was conducted on the modified IEEE 13 bus system from [8] and the modified IEEE 34 bus system from [29], as shown in Figure 8. In this study, we did not consider the size and the connected position of PV systems. Once the PV system installed by its owner is in the distribution system, the size and the connected position of the PV system cannot change at all. We focused on the operation strategy of the distribution system operator in the situation after the topology of the distribution network has already been determined. Therefore, the data about the topology of the test system is extracted from other references which are adopting PV inverters in distribution systems. In this section, the detailed experimental environment and a comparison with other control methods are presented.

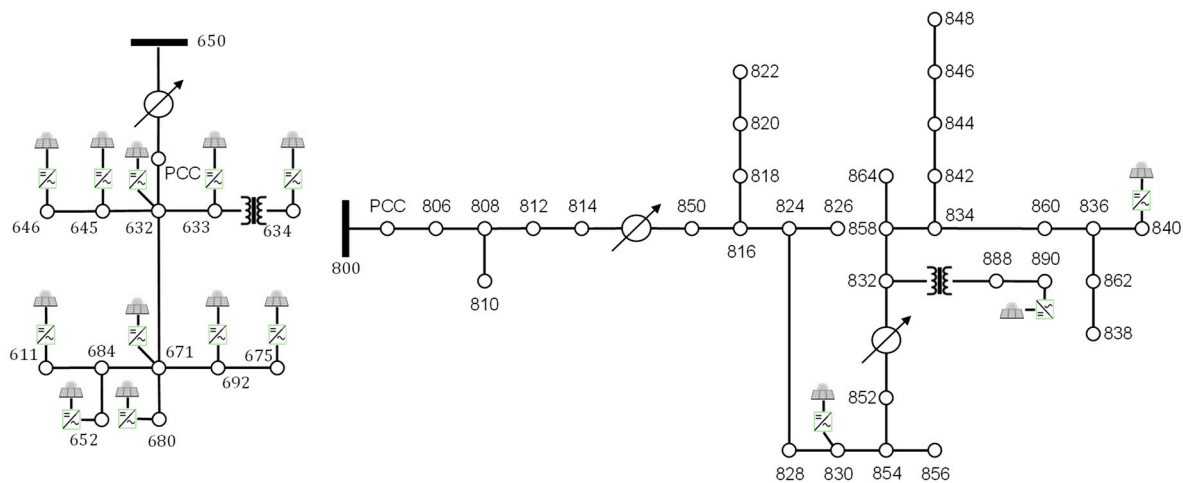


Figure 8. Modified IEEE 13 bus system (left) and IEEE 34 bus system (right) in a single-line diagram.

4.1. Simulation Data

The PV systems were added at buses 646, 645, 632, 633, 634, and 671 line-to-line and 611, 692, 675, 652, and 680 line-to-neutral in IEEE 13 bus system. In IEEE 34 bus system 3 phase PV systems are added at buses 830 line-to-neutral and 840 and 890 line-to-line. Table 1 shows the phase and rated active power of each PV system. The phases AB, BC, and CA are represented as 1, 2, and 3 for PV systems connected line-to-line, and A, B, and C are represented as 1, 2, and 3 for PV systems connected line-to-neutral.

Table 1. Identification of single-phase PV systems.

Network	Line-to-Line						Line-to-Neutral								
	IEEE 13 Bus			IEEE 34 Bus			IEEE 13 Bus								
Location	646	645	632	633	634	671	840	890	830	611	692	675	652	680	
Capacity	S_i^{rated} (kVA)	110	110	110	165	70	60	110	110	110	60	110	122	110	60
	P_i^{rated} (kW)	100	100	100	150	60	50	100	100	100	50	100	110	100	50
Phase	2	2	1	1	1	3	1, 2, 3	1, 2, 3	1, 2, 3	3	3	1	1	1	

4.2. Comparison

To verify the performance of the proposed strategy, three different cases were examined:

- Case 1: inverters with volt–var curve controlled by the proposed method
- Case 2: inverters with volt–var curve with a fixed value
- Case 3: inverters with no reactive power compensation

In case 1, V_1 , V_2 , V_3 , V_4 at each inverter volt–var curve were tuned by the proposed method. In case 2, the V_1 , V_2 , V_3 , V_4 at each inverter volt–var curve were 0.92, 0.98, 1.02, and 1.08, respectively. In case 3, inverters with no reactive power compensation were used as the base case for comparison with cases 1 and 2.

4.3. Simulation Result

Figure 9 shows the reward profile of each epoch, where the epochs represent the number of iterations. With epochs at 2000, the reward almost converges, which means that the training is almost complete; therefore, the variables of the neural nets are rarely updated. Thus, the policy that returns optimal action when observing the state has been trained.

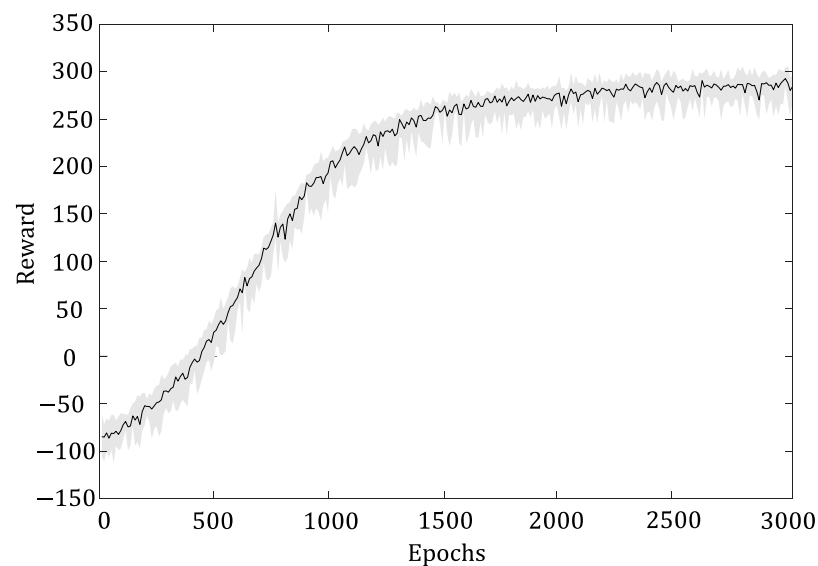


Figure 9. Reward profile.

After the convergence of the reward, the experiment was conducted on cloudy and sunny days to evaluate the performance of trained inverters. The Figures 10a and 11a shows the normalized PV generation and load consumption for each day, which are applied to various PV generation and load consumptions per unit. The Figures 10b and 11b shows the $VUF(\%)$ at the PCC in modified IEEE 13 bus system for each case during the day. The Figures 10c and 11c show the $VUF(\%)$ at the PCC in the modified IEEE 34 bus system for each case during the day. Table 2 shows the average $VUF(\%)$ at the PCC for each case during the day.

Table 2. Average $VUF(\%)$ for each case during both conditions.

		Network	Case 1	Case 2	Case 3
$VUF(\%)$	Sunny day	IEEE 13	1.23	1.38	1.37
		IEEE 34	0.05	0.0514	0.52
	Cloudy day	IEEE 13	1.29	1.40	1.42
		IEEE 34	0.053	0.054	0.055

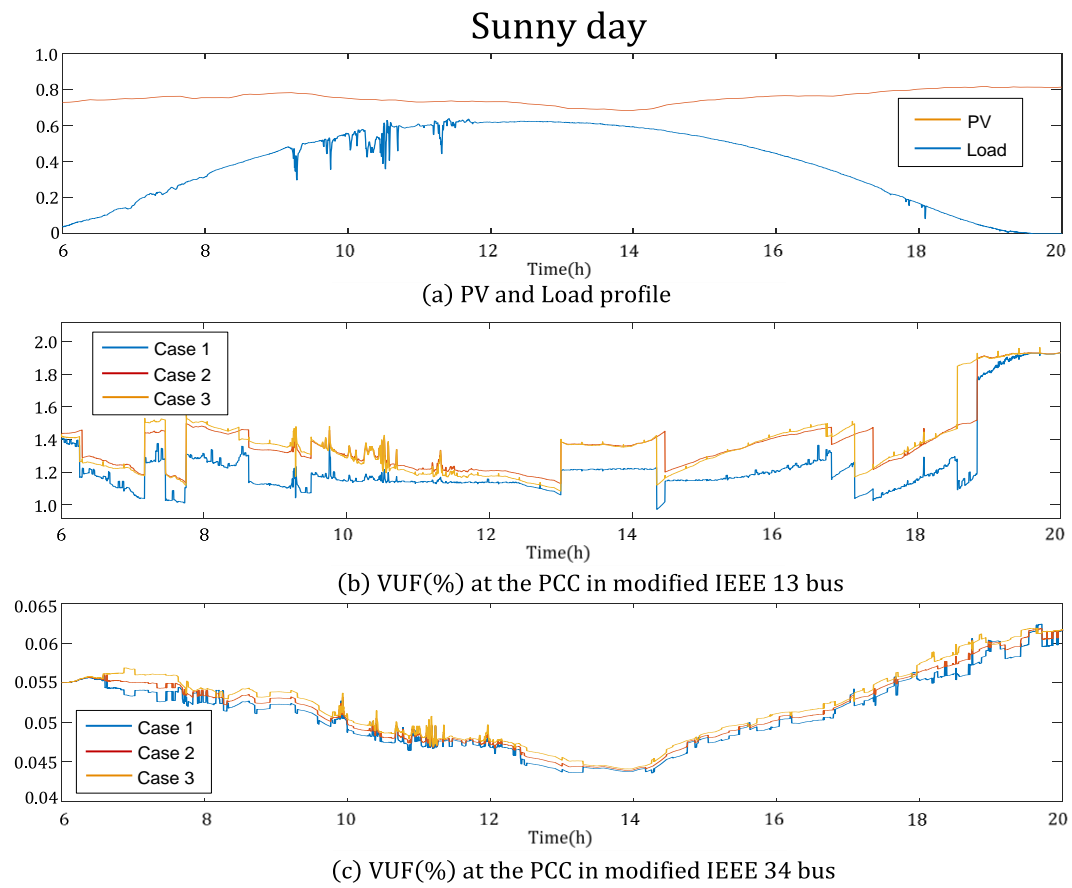


Figure 10. PV and load profile and $VUF(\%)$ at the PCC for a sunny day.

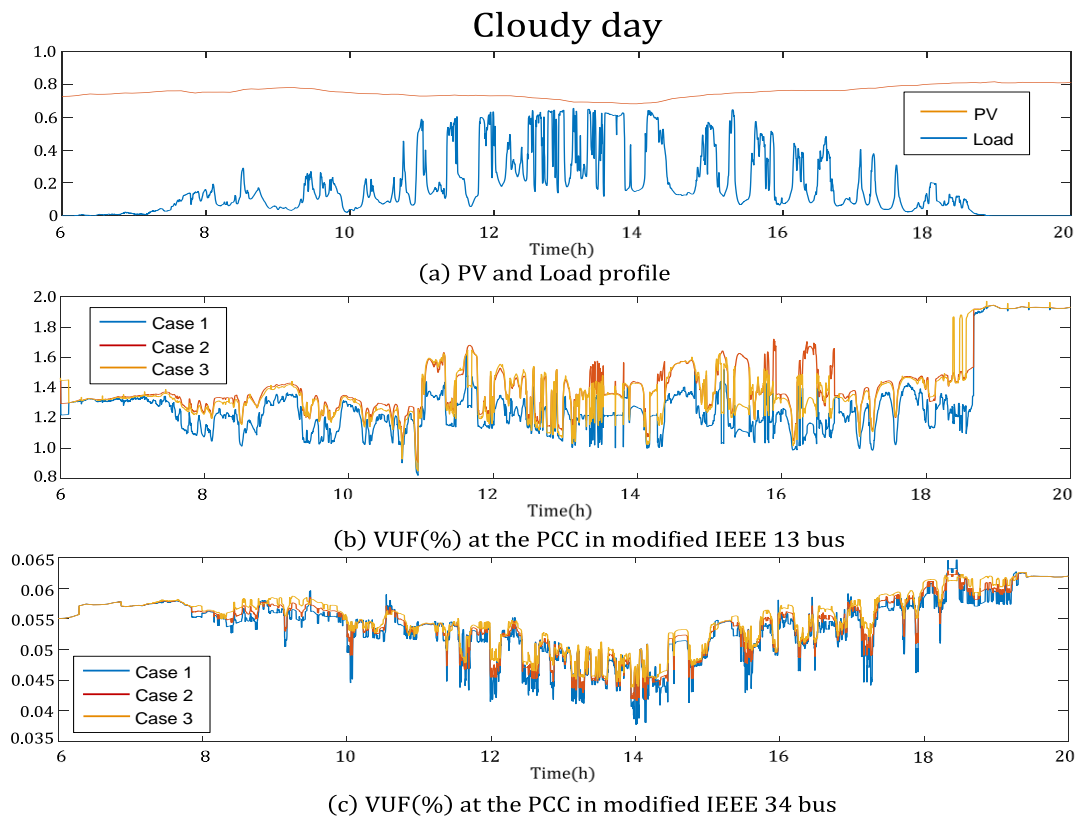


Figure 11. PV and load profile and $VUF(\%)$ at the PCC for a cloudy day.

In Figure 11, in the time period of 6~7 h and 19~20 h, there doesn't seem to be any difference between $VUF(\%)$ of each case. It is not allowed to provide/absorb reactive power from each PV inverter when the inverter's active power generation is less than $0.05 \times P_i^{rated}$. So the $VUFs(\%)$ of each case are the same while normalized PV is lower than 0.05. Additionally, the $VUF(\%)$ of case 1 is lower than others for each day. This means that the proposed strategy mitigates the voltage unbalance at the PCC more than for other comparisons. However, the proposed strategy compensates the reactive power based on its volt-var curve for each inverter and not the mathematically calculated value. Therefore, it is impossible to make the $VUF(\%)$ zero, but it is a significant advantage to compensate reactive power in real-time according to its volt-var curve. It is expected that the $VUF(\%)$ of case 1 is lower than that of case 2 for the sunny condition because of the similarity between the actual and forecast profiles of PV and load. Moreover, the outstanding performance of case 1 can be seen in cloudy conditions. This demonstrates that the proposed strategy can also act appropriately even when far from the forecast data owing to cloudiness and other factors. The effect of case 2 was either insignificant or made matters worse than case 3, despite the compensation of reactive power based on the fixed volt-var curve. In other words, inappropriate volt-var control should be discouraged.

5. Conclusions and Future Work

A novel method for mitigating voltage unbalance at the PCC by tuning the volt-var curve of each PV inverter through a day-ahead DRL training platform with forecast data in a digital twin grid was proposed. The simulation results demonstrated the performance of the proposed method. The auxiliary service of DG, such as volt-var control, becomes necessary because of the increasing proportion of DG in the power system. Moreover, the improvement of auxiliary service of DG leads to an increased capacity for renewable energy. Related future works include the further application of the proposed method to ensure transient voltage stability. Moreover, we will develop the strategy for the incentives of this auxiliary service to encourage participation by PV system owners.

Author Contributions: Conceptualization, C.H., G.J. and Y.J.; methodology, C.H. and Y.J.; software, C.H., D.L. and Y.J.; validation, C.H., S.S. and Y.J.; formal analysis, G.J., C.H. and S.S.; investigation, C.H. and Y.J.; resources, C.H. and Y.J.; data curation, C.H. and Y.J.; writing—original draft preparation, Y.J.; writing—review and editing, C.H.; visualization, C.H., D.L. and Y.J.; supervision, G.J. and Y.J.; project administration, C.H. and Y.J.; funding acquisition, C.H. and G.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MOTIE) (No. 20191210301890) and Korea Electric Power Corporation (Grant number: R20XO02-4).

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

References

1. Putrus, G.A.; Suwanapingkarl, P.; Johnston, D.; Bentley, E.C.; Narayana, M. Impact of electric vehicles on power distribution networks. In Proceedings of the 2009 IEEE Vehicle Power and Propulsion Conference, Dearborn, MI, USA, 7–10 September 2009; pp. 827–831.
2. Balamurugan, K.; Srinivasan, D.; Reindl, T. Impact of distributed generation on power distribution system. *Energy Procedia* **2012**, *25*, 93–100. [[CrossRef](#)]
3. Kersting, W.; Phillips, W. Phase frame analysis of the effects of voltage unbalance on induction machines. *IEEE Trans. Ind. Appl.* **1997**, *33*, 415–420. [[CrossRef](#)]
4. Lee, C.Y. Effects of unbalanced voltage on the operation performance of a three-phase induction motor. *IEEE Trans. Energy Convers.* **1999**, *14*, 202–208.
5. Jha, R.R.; Dubey, A.; Liu, C.C.; Schneider, K.P. Bi-level volt-var optimization to coordinate smart inverters with voltage control devices. *IEEE Trans. Power Syst.* **2019**, *34*, 1801–1813. [[CrossRef](#)]
6. Nájera, J.; Mendonça, H.; de Castro, R.M.; Arribas, J.R. Strategies comparison for voltage unbalance mitigation in LV distribution networks using EV chargers. *Electronics* **2019**, *8*, 289. [[CrossRef](#)]

7. Nejabatkhah, F.; Li, Y.W. Flexible unbalanced compensation of three-phase distribution system using single-phase distributed generation inverters. *IEEE Trans. Smart Grid* **2019**, *10*, 1845–1857. [[CrossRef](#)]
8. Yao, M.; Hiskens, I.A.; Mathieu, J.L. Mitigating Voltage Unbalance Using Distributed Solar Photovoltaic Inverters. *IEEE Trans. Power Syst.* **2021**, *36*, 2642–2651.
9. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)]
10. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [[CrossRef](#)]
11. Duan, J.; Yi, Z.; Shi, D.; Lin, C.; Lu, X.; Wang, Z. Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid AC–DC microgrids. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5355–5364. [[CrossRef](#)]
12. Duan, J.; Shi, D.; Diao, R.; Li, H.; Wang, Z.; Zhang, B.; Bian, D.; Yi, J. Deep-reinforcement-learning-based autonomous voltage control for power grid operations. *IEEE Trans. Power Syst.* **2019**, *35*, 814–817. [[CrossRef](#)]
13. Wang, W.; Yu, N.; Shi, J.; Gao, Y. Volt-VAR control in power distribution systems with deep reinforcement learning. In Proceedings of the 2019 IEEE International Conference on Communications, Control and Computing Technologies for Smart Grids (SmartGridComm), Beijing, China, 21–23 October 2019; pp. 1–7.
14. Huang, Q.; Huang, R.; Hao, W.; Tan, J.; Fan, R.; Huang, Z. Adaptive power system emergency control using deep reinforcement learning. *IEEE Trans. Smart Grid* **2020**, *11*, 1171–1182. [[CrossRef](#)]
15. Zhang, Y.; Wang, X.; Wang, J.; Zhang, Y. Deep reinforcement learning based volt-var optimization in smart distribution systems. *IEEE Trans. Smart Grid* **2020**, *12*, 361–371. [[CrossRef](#)]
16. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
17. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital twin: Enabling technologies, challenges and open research. *IEEE Access* **2020**, *8*, 108952–108971. [[CrossRef](#)]
18. Jouanne, A.V.; Banerjee, B. Assessment of voltage unbalance. *IEEE Trans. Power Deliv.* **2001**, *16*, 782–790. [[CrossRef](#)]
19. Singh, A.K.; Singh, G.K.; Mitra, R. Some observations on definitions of voltage unbalance. In Proceedings of the 2007 39th North American Power Symposium, Las Cruces, NM, USA, 30 September–2 October 2007; pp. 473–479.
20. Dzafic, I.; Donlagic, T.; Henselmeyer, S. Fortescue transformations for three-phase power flow analysis in distribution networks. In Proceedings of the 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22–26 July 2012; pp. 1–7.
21. IEEE standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces. In *IEEE Std 1547–2018 (Revis. IEEE Std 1547–2003)*; IEEE: Piscataway Township, NJ, USA, 2018; pp. 1–138.
22. Lipowski, A.; Lipowska, D. Roulette-wheel selection via stochastic acceptance. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 2193–2196. [[CrossRef](#)]
23. Korea Power Exchange. Available online: <http://kpx.or.kr> (accessed on 16 August 2021).
24. Tostado-Véliz, M.; Kamel, S.; Jurado, F. Development and Comparison of Efficient Newton-Like Methods for Voltage Stability Assessment. *Electr. Power Compon. Syst.* **2020**, *48*, 1798–1813.
25. Swief, R.A.; Hassan, N.M.; Abdelaziz, A.Y.; Kamh, M.Z. Multi-Regional Optimal Power Flow Using Marine Predators Algorithm Considering Load and Generation Variability. *IEEE Access* **2021**, *9*, 74600–74613. [[CrossRef](#)]
26. Tostado-Véliz, M.; Kamel, S.; Jurado, F. Power flow solution of Ill-conditioned systems using current injection formulation: Analysis and a novel method. *Int. J. Electr. Power Energy Syst.* **2021**, *127*, 106669. [[CrossRef](#)]
27. Dugan, R.C.; McDermott, T.E. An open source platform for collaborating on smart grid research. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 24–28 July 2011; pp. 1–7.
28. PyTorch. Available online: <http://pytorch.org> (accessed on 19 August 2021).
29. Bedawy, A.; Yorino, N.; Mahmoud, K.; Zoka, Y.; Sasaki, Y. Optimal Voltage Control Strategy for Voltage Regulators in Active Unbalanced Distribution Systems Using Multi-Agents. *IEEE Trans. Power Syst.* **2019**, *35*, 1023–1035. [[CrossRef](#)]