Graph Reinforcement Learning for Volt/VAR Control in Three-Phase Unbalanced Distribution Networks Using a U-Net Structure

Fan Huang*, Cuo Zhang
School of Electrical and Computer Engineering
The University of Sydney, Australia
*Email: fan.huang@sydney.edu.au

Xiao Liu
Faculty of Environment, Science and Economy
University of Exeter, United Kingdom
Email: X.Liu8@exeter.ac.uk

***Abstract*—Data-driven Volt/VAR Control (VVC) methods provide an effective means to enhance bus voltage regulation of distribution networks and thus mitigate photovoltaic (PV) curtailment under high penetration. However, their extension to three-phase unbalanced distribution networks remains underexplored, where strong phase coupling and radial topology introduce additional coordination challenges. Graph Reinforcement Learning (GRL) has recently emerged as a promising approach for VVC by leveraging the underlying network topology. Nevertheless, conventional shallow architectures struggle to capture voltage dependencies among distant buses and to coordinate decisions across the entire system. To address these challenges and minimize PV curtailment, this paper proposes a U-shaped Graph Neural Network (UGNN) for inverter-based VVC, trained via the soft actor–critic algorithm. UGNN captures hierarchical feature dependencies and global coordination through a U-Net–like pooling/unpooling structure. Node selection during pooling is guided by Laplacian positional encoding and node feature inputs. During unpooling, hybrid residual paths preserve both global context and local voltage details. Evaluations on three-phase unbalanced distribution networks demonstrate that the UGNN achieves markedly faster early-stage convergence, efficiently discovering feasible and near-optimal control strategies compared with existing GNN-based VVC methods.**

***Index Terms*—Graph neural network, PV curtailment minimization, reinforcement learning, unbalanced distribution network, Volt/VAR control.**

I. Introduction

The global transition toward clean and sustainable energy has driven the rapid deployment of distributed photovoltaic (PV) generation in distribution networks. However, the high penetration of PV introduces significant operational challenges, particularly bus voltage rise and further leading to substantial PV power generation curtailment [1]–[3]. Such PV curtailment, though simple to implement for mitigating overvoltage issues, is economically inefficient and undermines the sustainability of renewable integration by reducing energy yield and profitability for PV owners [3]. Therefore, developing efficient PV inverter-based Volt/VAR Control (VVC) methods has become essential for maintaining voltage profiles and supporting large-scale renewable integration [4], [5].

Traditional VVC methods, including model-based centralized optimization, rule-based local control, and hierarchical coordination of PV inverters, can provide voltage regulation and control, and minimize network power loss in distribution networks. However, these methods work well for single-phase or equivalent distribution networks and struggle to handle the PV curtailment minimization problems with an accurate network model and under strong renewable uncertainty. Recent studies proposed linearized network operation formulations to improve scalability and reduce computational burdens [6]. Even so, their performance remains sensitive to network and control modeling complexity, system size, and requirement on solution time. These limitations highlight the need for data-driven VVC methods that can efficiently adapt to dynamic operating conditions in three-phase unbalanced distribution networks.

Data-driven VVC methods have recently attracted increasing attention as they can learn complex operational patterns directly from data without relying on explicit physical models. Among them, reinforcement learning (RL) methods have emerged as a powerful framework for handling nonlinear system dynamics and uncertainty in unbalanced distribution networks [7]–[9]. Representative studies have applied both on-policy and off-policy RL algorithms such as Soft Actor–Critic (SAC) to coordinate inverter reactive power outputs, demonstrating superior adaptability and decision efficiency compared with conventional model-based methods [8], [9]. Building on these advances, recent work has begun to integrate graph neural networks (GNNs) into RL algorithms, enabling the learning process to explicitly utilize network topology and spatial correlations among buses in unbalanced distribution networks [10], [11].

Recent studies on Graph Reinforcement Learning (GRL) have explored data-driven VVC to coordinate inverter actions and minimize PV curtailment in distribution networks [10]. These methods leverage network topology to coordinate inverter reactive power and curtailment decisions, achieving improved voltage regulation and operational efficiency compared with conventional model-based controllers. Nevertheless, most of these graph RL applications still assume shallow message-passing layers, which restrict their receptive field and limit their ability to capture long-range voltage relationship across

979-8-3315-8138-1

long feeders [10]. In addition, deeper GNNs or multi-hop architectures introduce substantial computational overhead, making them challenging for real-time deployment in unbalanced distribution networks [10], [11]. These observations suggest that, while topology-aware GNNs hold promise, further investigation is imperative into structure that explicitly model three-phase, system-wide coupling for VVC under high PV uncertainty.

To address the limited receptive field and lack of global coordination in existing GNN-based VVC methods, this paper proposes a *U-shaped Graph Neural Network (UGNN)* for minimizing PV curtailment in three-phase unbalanced distribution networks. The UGNN adopts a hierarchical encoder–decoder structure that aggregates global voltage features and restores local information with physical consistency, enabling coordinated control across distant buses. Experimental results demonstrate that the proposed method improves voltage regulation and convergence speed while maintaining stable performance under varying operating conditions.

II. Formulation of Volt/VAR Control as a Markov Decision Process

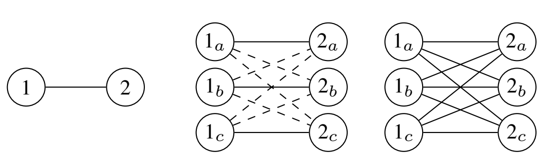
This study formulates the inverter-based VVC problem for three-phase unbalanced distribution networks as a Markov Decision Process (MDP), where the control policy aims to minimize PV curtailment while maintaining all phase voltages within their operational limits. The distribution network is modeled by a three-phase unbalanced current-injection formulation solved via a backward/forward sweep (BFS) power-flow algorithm, which defines the physical state transition between decision steps.

1) Network Representation

Let $\mathcal{G}_{\text{bus}} = (\mathcal{V}_{\text{bus}}, \mathcal{E}_{\text{bus}})$ be the bus-level network, where B denotes the number of buses and $A_{\text{bus}} \in \{0,1\}^{B\times B}$ is a sparse adjacency matrix without self-loops. To reflect three-phase coupling, a phase-coupled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is constructed with $\mathcal{V} = \mathcal{V}_{\text{bus}} \times \{a, b, c\}$. The adjacency $A \in \{0,1\}^{N\times N}$ is defined as

$$A \;=\; A_{\text{bus}} \otimes J_3, \tag{1}$$

where $J_3 = \mathbf{1}_3\mathbf{1}_3^\top$ is the phase-coupling expansion matrix and $N = 3B$ is the total number of phase nodes. This construction connects all phase pairs for each bus pair, as illustrated in Fig. 1.



(a) Branch in $\mathcal{G}_{\text{bus}}$ (b) Phase-expanded (from $A_{\text{bus}}\otimes I_3$) (c) Phase-coupled (from $A_{\text{bus}}\otimes J_3$)

Fig. 1. Illustration of bus-level and phase-coupled graphs used for three-phase modeling.

2) State Space

At each control step t, the system state s_t consists of phase-level per-unit voltages and normalized global power descriptors, i.e.,

$$s_t = (X_t,\, g_t), \tag{2}$$

where $X_t = [\, V_{n,z} \,]_{(n,z)\in\mathcal{V}} \in \mathbb{R}^N$ denotes the vector of phase voltages, and $g_t = [\, P_\Sigma^{\text{load}},\, Q_\Sigma^{\text{load}},\, P_\Sigma^{\text{PV}} \,] \in \mathbb{R}^3$ collects global active/reactive power descriptors.

3) Action Space

For each controllable PV inverter, the action comprises the curtailed active power and reactive power set-point,

$$\mathbf{a}_t = [\, P_{n,z}^{\text{curt}},\, Q_{n,z}^{\text{PV}} \,]_{(n,z)\in\mathcal{N}_{\text{PV}}}. \tag{3}$$

The dispatched active power is calculated as $P_{n,z}^{\text{PV}} = P_{n,z}^{\text{PV,avail}} - P_{n,z}^{\text{curt}}$, subject to the apparent-power constraint

$$\big(P_{n,z}^{\text{PV}}\big)^2 + \big(Q_{n,z}^{\text{PV}}\big)^2 \le \big(S_{n,z}^{\text{PV}}\big)^2, \quad \forall (n,z) \in \mathcal{N}_{\text{PV}}. \tag{4}$$

Positive $Q_{n,z}^{\text{PV}}$ indicates reactive injection, while negative values indicate absorption.

4) State Transition

The environment evolves deterministically according to the three-phase BFS power flow [12], which updates nodal voltages as

$$\mathbf{V}_{t+1} = \text{BFS}\big(s_t, \mathbf{a}_t\big), \tag{5}$$

yielding the next state $s_{t+1} = (X_{t+1}, g_{t+1})$.

5) Reward Function

The reward penalizes voltage violations and PV curtailment:

$$\begin{aligned} r = -\,\lambda_V \Bigg[\sum_{n,z} \Big([V_{n,z} - V_{\max}]_+^2 + [V_{\min} - V_{n,z}]_+^2 \Big) \\ +\, \eta\, N_{\text{viol}} \Bigg] - \lambda_C \sum_{n,z} P_{n,z}^{\text{curt}}. \end{aligned} \tag{6}$$

Here $[x]_+ = \max(x, 0)$ extracts the over-limit portion of voltage magnitude, N_{viol} counts the number of buses violating $[V_{\min}, V_{\max}]$, and λ_V, λ_C, and η are weighting coefficients for voltage, curtailment, and violation penalties, respectively. This design encourages the policy to keep voltages within prescribed limits while minimizing unnecessary curtailment.

6) Objective

The agent seeks a policy $\pi_\theta(\mathbf{a}_t \,|\, s_t)$ that maximizes the discounted return

$$\max_{\pi_\theta}\; \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t \right],$$

with discount factor $\gamma \in (0, 1)$ and horizon T.

979-8-3315-8138-1

III. U-Shaped Graph Reinforcement Learning

Realistic three-phase unbalanced distribution networks exhibit stronger phase coupling and more diverse phase-level characteristics than a single-phase approximation. To capture such inter-phase dependencies, a phase-coupled graph expansion is combined with a U-shaped hierarchical mechanism that maps fine-grained phase-level voltages to progressively abstract system-level representations and subsequently restores information back to the phase level for coordinated control.

Figure 2 depicts an inverted U-shape overview with pooling on the upper path and unpooling on the lower path. This section introduces the overall architecture at a conceptual level, while detailed formulations and implementation aspects are presented in the subsequent subsections.

A. Normalized Propagation on Phase-Coupled Graphs

At hierarchical level k, corresponding to the k-th graph scale in the U-shaped architecture, let A_k denote the adjacency matrix of the induced subgraph after the previous pooling stage, and $H^{l,k} \in \mathbb{R}^{N_k \times d_l}$ the hidden representations of its N_k nodes at layer l. Define the augmented adjacency $\tilde{A}_k = A_k + I$, and let $\tilde{D}_k$ be the degree matrix with entries $(\tilde{D}_k)_{ii} = \sum_j (\tilde{A}_k)_{ij}$. One propagation step is

$$H^{l+1,k} = \sigma\Big[\tilde{D}_k^{-1/2}\,\tilde{A}_k\,\tilde{D}_k^{-1/2}\,H^{l,k}\,W^{l,k}\Big], \tag{7}$$

where $W^{l,k}$ are trainable weights and σ is a nonlinearity. A small fixed number of such steps is applied before each pooling stage.

B. Hierarchical Region Representative Selection

Pooling selects region representatives and forms level-wise regions, corresponding to the upper path in Fig. 2. Positional information is encoded using Laplacian positional encodings (LapPE) computed from the first d_{pe} Laplacian eigenvectors [13]. At level k with N_k nodes, let $P^k \in \mathbb{R}^{N_k \times d_{\text{pe}}}$ denote the precomputed LapPE with retained dimension d_{pe}. Let $H^{l,k} \in \mathbb{R}^{N_k \times d_l}$ be the current node features. Here N_k denotes the number of nodes at level k, and d_l the feature dimension at the current layer.

A pooling ratio $\rho_k \in (0,1)$ determines the number of representatives retained at the next level, given by $N_{k+1} = \lceil \rho_k N_k \rceil$. Based on N_{k+1}, a learned soft assignment matrix $S_k = f_{\text{pool}}\big[H^{l,k} \,\|\, P^k\big] \in \mathbb{R}^{N_k \times N_{k+1}}$ is produced by a small Multi-Layer Perceptron (MLP) followed by row-wise softmax with temperature τ_k that controls selection sharpness. Each row of S_k gives the normalized weights from a node to the N_{k+1} representatives and sums to one.

Representative features are then aggregated as

$$Z^{k+1} = S_k^{\top} H^{l,k} \in \mathbb{R}^{N_{k+1} \times d_l}, \tag{8}$$

where $Z^{(k)}$ denotes the representative feature matrix at level k and $d^{(k)}$ its feature dimensionality. As indicated by (8), the resulting Z^{k+1} summarizes the member buses assigned to each representative. Propagation continues at level $k+1$ on the induced meta-graph using the normalized update from the previous subsection. LapPE P^k stabilizes scoring and promotes spatially coherent, physically plausible regions.

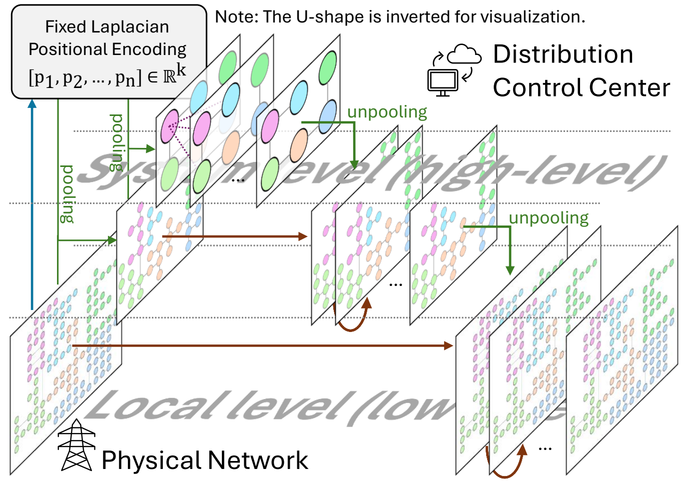


Fig. 2. Inverted U-shape overview with pooling, unpooling, and LapPE.

C. Global Coordination and Local Detail Recovery

During the upward pooling, phase-level details are progressively coarsened. After reaching the top level, UGNN propagates coarse features to capture global context and facilitate detail restoration.

Top-level propagation with global context. At the top level K, let $g \in \mathbb{R}^{d_g}$ denote the global summary. Before each propagation, g is concatenated once to every node feature at this level, yielding

$$H_{\text{aug}}^{l,K} = \big[\,H^{l,K}\,;\,\mathbf{1}g^{\top}\big] \in \mathbb{R}^{N_K \times (d_l + d_g)}, \tag{9}$$

that is, the features are concatenated and the dimension increases from d_l to $d_l + d_g$. The normalized propagation then follows Eq. (7) on $H_{\text{aug}}^{l,K}$.

Downward restoration via unpooling. After coarse propagation, restoration proceeds level by level. Unpooling follows the reverse of (8), with the learned assignment S_k distributing level-$k{+}1$ representative features back to nodes. The decoder at level k is initialized by a single residual bridge that merges encoder skips with lifted coarse features:

$$H^{0,k} \;=\; \alpha_k\,H_{\text{skip}}^{k} \;+\; (1-\alpha_k)\,S_k\,Z^{k+1}, \tag{10}$$

Here H_{skip}^{k} denotes encoder skip features cached at level k and aligned to d_l, the term $S_k Z^{k+1} \in \mathbb{R}^{N_k \times d_l}$ provides coarse guidance from representatives and $\alpha_k \in [0,1]$ preserves fine details through skips.

Then refined using the normalized propagation from (7) together with a light layerwise residual that sustains coarse guidance:

$$H^{l+1,k} \;=\; \text{GNN}^{l,k}\big(H^{l,k}\big) \;+\; \beta_k\,S_k\,Z^{k+1}, \tag{11}$$

Here $\beta_k \in (0,1]$ keeps the guidance light so it does not overwhelm restored details.

By the final level, representative information has been passed down and combined with local features, yielding phase-level results that keep both detail and system consistency.

D. Training Procedure with SAC

UGNN serves as a shared topology-aware encoder for the actor and critics in a SAC framework. All remaining training

Algorithm 1 UGNN–SAC Training with Prioritized Experience Replay (PER) and Entropy Cap

Require: Graph $\mathcal{G}$, adjacency A, cached Laplacian PEs $\{P^k\}$ for pooling, pooling ratios $\{\rho_k\}$, discount γ, target entropy $\bar{\mathcal{H}}$, learning rates $\eta_{\text{actor}}, \eta_{\text{critic}}, \eta_{\text{temp}}$, Polyak rate τ, horizon T, replay buffer $\mathcal{M}$ with warmup $N_{\min}$
Ensure: Policy π_ϕ, critics Q_{ψ_1}, Q_{ψ_2}

```
Initialize shared UGNN encoder f_θ and cache {P^k}; initialize actor π_φ, critics Q_ψ1, Q_ψ2 and target networks Q'_ψ1, Q'_ψ2; initialize temperature α.
for t = 1 to T do
  Observe state s_t; generate action a_t with π_φ; apply BFS_POWERFLOW to obtain (s_{t+1}, r_t).
  Store (s_t, a_t, r_t, s_{t+1}) into replay buffer M.
  if |M| ≥ N_min then
    Sample a minibatch from M using PER.
    Compute target Q values and the entropy term ent_term.
    Check whether ent_term exceeds a cap defined as a fixed ratio of the target Q value.
    if not exceeded then
      Update actor, critics, and temperature α.
    else
      Clip ent_term to the cap; update actor and critics with the clipped term; stop updating α.
    end if
    Update target critics by Polyak averaging.
  end if
end for
Note: Cached Laplacian PEs {P^k} are used for pooling score computation in the UGNN encoder.
```

details and calls outside the UGNN internals are summarized in the pseudocode below.

Remark. In the late stage of VVC training, when the critic Q magnitude decreases, cap the entropy contribution once its ratio to the Q term exceeds a threshold, and then freeze α. This stabilization rule markedly improves convergence stability and final performance for a wide range of networks using SAC on the VVC task.

IV. Case Study

A. Test System Description

A modified IEEE 123-bus test system with PV installations is adopted, where the system parameters are taken from [14]. The PV capacity is scaled by a factor of two to emulate a high-penetration scenario. A three-phase unbalanced distribution network is simulated using a BFS power flow solver to ensure physically consistent state transitions between control steps.

Four agents are evaluated: MLP, GCN, UGNN-1, and UGNN-2. UGNN-1 includes one pooling level, while UGNN-2 includes two. All GNN-based models are employed as feature decoders to ensure architectural fairness in network functionality, while the MLP agent directly maps input features to an output space of the same dimensionality as the GNN decoders. At each control step, loads and PV injections are perturbed by Gaussian multipliers clipped within nominal ranges to represent stochastic operating conditions.

TABLE I
Network configuration of shared GNN encoder and SAC heads.

Shared GNN			
Component	GCN	UGNN-1	UGNN-2
Hidden dims per node	16	16	16
Hidden global dims	16	16	16
Output dim	16	16	16
Dropout	0.1	0.1	0.1
Number of pools	0	1	2
Pool ratio	–	0.25	0.25
LapPE enabled	–	True	True
Residual $(\alpha, \alpha_g, \beta)$	–	(1.0, 0.3, 0.5)	(1.0, 0.3, 0.5)
SAC Head			
	Actor	**Critic**	
Input feature	Shared GNN output	Shared GNN output + action	
Hidden dims	[128, 64]	[128, 64]	
Dropout	0.1	0.1	

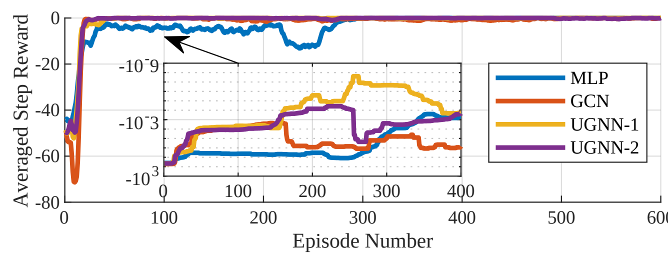


Fig. 3. Averaged step reward per episode over time.

Since VVC is a quasi-static control task, the discount factor is set to zero to emphasize immediate voltage regulation. The MLP agent relies heavily on replay-driven updates, and a small replay buffer of 5,000 samples is used to maintain update freshness. Each episode consists of 50 control steps, and the batch size is 128. All agents share identical SAC optimizers and training configurations.

B. Ablation Study

Figure 3 shows the reward evolution over training, where UGNN-1, UGNN-2, and GCN all converge rapidly, while MLP exhibits a noticeably slower learning rate. Figure 4 compares the voltage regulation performance; although GCN also converges quickly, it presents larger voltage fluctuations than the UGNN models, indicating less stable voltage coordination. As shown in Figure 5, both UGNN variants achieve smoother and more favorable PV curtailment trajectories relative to GCN.

Table II summarizes the area-under-curve (AUC) results across different training stages. Overall, UGNN architectures exhibit comparable or slightly better sample efficiency than GCN, while MLP trails behind. Between the two UGNN configurations, differences are task-dependent: the shallower and deeper variants show complementary strengths across metrics rather than a strict dominance relationship.

It is also observed that UGNNs are extremely sensitive to hyperparameter tuning. With a reliable adaptive tuning strategy, their performance could be further enhanced.

979-8-3315-8138-1

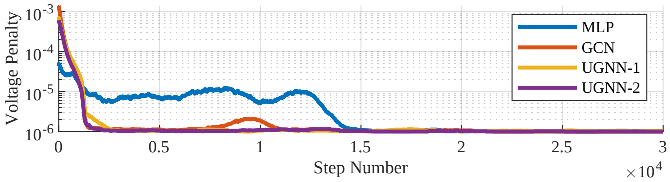


Fig. 4. Voltage penalty over training steps.

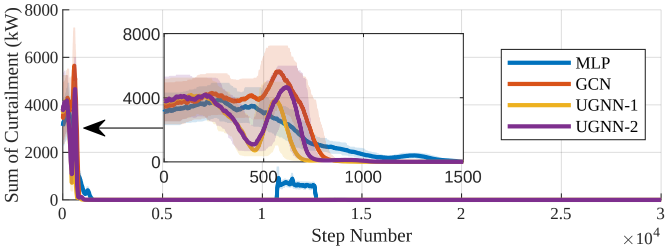


Fig. 5. Total PV curtailment over training steps.

TABLE II
COMPARISON OF AUC ACROSS DIFFERENT TRAINING STAGES.

Agent	Ep	Step reward	Voltage penalty	Curtailment(kW)
MLP	50	7.98×10^2	1.38×10^4	-2.60×10^6
	200	1.49×10^3	4.65×10^4	-2.60×10^6
	400	2.10×10^3	6.22×10^4	-3.67×10^6
GCN	50	9.11×10^2	1.51×10^4	-3.06×10^6
	200	9.63×10^2	1.80×10^4	-3.07×10^6
	400	1.09×10^3	1.87×10^4	-3.07×10^6
UGNN-1	50	7.01×10^2	1.56×10^4	$\mathbf{-1.98 \times 10^6}$
	200	7.09×10^2	1.59×10^4	$\mathbf{-1.98 \times 10^6}$
	400	$\mathbf{7.25 \times 10^2}$	1.66×10^4	$\mathbf{-1.98 \times 10^6}$
UGNN-2	50	$\mathbf{6.97 \times 10^2}$	$\mathbf{1.20 \times 10^4}$	-2.32×10^6
	200	$\mathbf{7.06 \times 10^2}$	$\mathbf{1.24 \times 10^4}$	-2.32×10^6
	400	$\mathbf{7.25 \times 10^2}$	$\mathbf{1.29 \times 10^4}$	-2.33×10^6

C. Generalization Across Power Operation Scenarios

Closed-loop Monte Carlo evaluations were conducted with 10,000 scenarios, where the active and reactive loads as well as PV generation were independently scaled by normalized factors uniformly sampled from $[0, 1]$ and applied to their nominal profiles. Since the training process used Gaussian perturbations around nominal conditions, these samples include out-of-distribution operating points not encountered during training.

Table III compares the best-history and best-recent checkpoints of each agent. UGNN-1 achieves the most robust historical model, maintaining perfect voltage compliance and strong stability under unseen conditions. UGNN-2 performs better in its most recent checkpoint, indicating higher latent learning capacity and continued improvement potential. In contrast, the GCN baseline shows inferior generalization compared with both UGNN variants.

V. CONCLUSION

This paper proposed a data-driven VVC framework for three-phase unbalanced distribution networks based on a UGNN integrated with the SAC algorithm. The proposed UGNN architecture aims to coordinate local fine-grained and global coarse-grained representations through a hierarchical design, addressing the intrinsic message-passing limitation of conventional GNNs. Integrated within a GRL framework for inverter-based VVC, it effectively mitigates PV curtailment and achieves robust voltage regulation across stochastic operating conditions, demonstrating clear advantages over standard GCN and MLP structures.

TABLE III
PERFORMANCE EVALUATION OF BEST MODELS UNDER POWER OPERATION SCENARIOS.

		$V_{\text{ref}} = 1.00$		$V_{\text{ref}} = 1.05$	
Agent	Type	Viol. (steps)	Avg Rwd	Viol. (steps)	Avg Rwd
GCN	Hist.	87	-1.017	87	-0.903
GCN	Rec.	106	-0.574	107	-0.567
UGNN-1	Hist.	**0**	**0.000**	**0**	**0.000**
UGNN-1	Rec.	13	-0.262	13	-0.286
UGNN-2	Hist.	10	-0.029	4	-0.007
UGNN-2	Rec.	**10**	**-0.029**	**4**	**-0.007**

ACKNOWLEDGMENT

This work was supported in part by Australian Research Council under Grant DE240100059 and The University of Sydney International Scholarship (USYDIS).

REFERENCES

[1] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.
[2] H. Ahmadi, J. R. Marti, and H. W. Dommel, "A linear power flow formulation for three-phase distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 5012–5021, 2016.
[3] Y. Z. Gerdroodbari, R. Razzaghi, and F. Shahnia, "Decentralized control strategy to improve fairness in active power curtailment of pv inverters in low-voltage distribution networks," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2282–2292, 2021.
[4] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Photovoltaic inverter controllers seeking ac optimal power flow solutions," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2809–2823, 2016.
[5] H. Soleimani Bidgoli and T. Van Cutsem, "Combined local and centralized voltage control in active distribution networks," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1374–1384, 2018.
[6] Q. Liang, C. Zhang, and J. Zhu, "Robust and risk-averse hosting capacity maximization for an unbalanced distribution network with volt/var control," *IEEE Transactions on Power Systems*, 2025, early Access.
[7] X. Liu, C. Zhang, S. Li, and J. Zhu, "A tuning friendly deep reinforcement learning method for inverter-based volt/var control," *IEEE Transactions on Smart Grid*, pp. 1–1, 2025.
[8] W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe off-policy deep reinforcement learning algorithm for volt–var control in power distribution systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3008–3018, 2020.
[9] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, and Z. Chen, "Deep reinforcement learning enabled physical-model-free two-timescale voltage control for active distribution systems," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 149–165, 2022.
[10] X. Y. Lee, S. Sarkar, and Y. Wang, "A graph policy network approach for volt–var control in power distribution systems," *Applied Energy*, vol. 323, p. 119530, 2022.
[11] D. Hu *et al.*, "Multi-agent graph reinforcement learning for decentralized volt–var control under partial observability," *Applied Energy*, vol. 363, p. 123401, 2024.
[12] J.-H. Teng, "A direct approach for distribution system load flow solutions," *IEEE Transactions on Power Delivery*, vol. 18, no. 3, pp. 882–887, 2003.
[13] V. P. Dwivedi, C. K. Joshi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson, "Benchmarking graph neural networks," *Journal of Machine Learning Research*, vol. 24, no. 43, pp. 1–48, 2023.
[14] Q. Liang, C. Zhang, and Z. Yang Dong, "Multi-objective volt/var control of virtual power plant for mitigating voltage unbalance," in *2022 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2022, pp. 419–423.

979-8-3315-8138-1