A Cross-Modal Generative Adversarial Network for Scenarios Generation of Renewable Energy

Mingyu Kang, Ran Zhu, Duxin Chen, Chaojie Li, *Member, IEEE*, Wei Gu, *Senior Member, IEEE*, Xusheng Qian, and Wenwu Yu, *Senior Member, IEEE*

***Abstract*—Scenarios data of renewable energy resources plays an essential role in the study of mitigating the risk in the power system due to their intermittent nature. Existing researches have mainly focused on how to generate high-quality time series based scenarios with corrupted or missing observed data. However, the multimodality of renewable energy resources is largely ignored from the perspective of the spatial-temporal correlation which can significantly strengthen the renewable energy prediction and optimization. To make full use of these multi-modal data and further improve the data quality of scenarios, a cross-modal Generative Adversarial Network (cGAN) model is proposed with the setting of two spatial-temporal transformer models. In cGAN, the task of scenarios generation is formulated as a probability approximation problem. Then, a cross-modal scenarios generation framework is constructed to process data from multi-modal observations. Theoretically, it can generate infinite number of uncertain scenarios data via the repeated random noise sampling technique. Extensive experiments are carried out at a NREL photovoltaic power dataset and a real-world wind power dataset, respectively. The results validate the state-of-the-art performance of cGAN, even though the validation dataset is missing at random. Moreover, the results also show that the setting of spatial-temporal transformer clearly explains the contributions of different modal data, and stabilize the training process of cGAN. Finally, a stochastic day-head economic dispatch is also studied to show the practical value of cGAN.**

Manuscript received 24 June 2022; revised 25 November 2022 and 19 February 2023; accepted 7 May 2023. Date of publication 18 May 2023; date of current version 21 February 2024. This work was supported in part by the National Key R&D Program of China under Grant 2022ZD0120004, in part by the National Natural Science Foundation of China under Grant 62233004, Grant 62273090, and Grant 62073076, in part by the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence under Grant BM2017002, and in part by the Australian Research Council under Grant DE210100274, in part by the Fundamental Research Funds for the Central Universities under Grant 2242023K40010. Paper no. TPWRS-00940-2022. *(Corresponding authors: Duxin Chen; Wenwu Yu.)*

Mingyu Kang and Ran Zhu are with the School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: kangmingyu.china@gmail.com; gemina_cat@163.com).

Duxin Chen is with the School of Mathematics, Southeast University, Nanjing 210096, China (e-mail: chendx@seu.edu.cn).

Chaojie Li is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Kensington, NSW 2052, Australia (e-mail: cjlee.cqu@163.com).

Wei Gu is with the School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: wgu@seu.edu.cn).

Xusheng Qian is with the Marketing Service Center (Metrology Center), State Grid Jiangsu Electric Power Company, Ltd., Nanjing 210019, China (e-mail: njqxsh@sina.com).

Wenwu Yu is with the Frontiers Science Center for Mobile Information Communication and Security, School of Mathematics, Southeast University, Nanjing 210096, China, and also with the Purple Mountain Laboratories, Nanjing 211102, China (e-mail: wwyu@seu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2023.3277698.

Digital Object Identifier 10.1109/TPWRS.2023.3277698

***Index Terms*—Scenarios generation, renewable energy, cross-modal generative adversarial network, photovoltaic power, wind power.**

I. Introduction

With the dramatically growing penetration of renewable energy resources, the dispatchability of photovoltaic (PV) and wind power have drawn much attentions where the intermittent nature of renewables could bring significant challenges to the operation and planning of modern power system. Importantly, to tackle the uncertainty, both stochastic and robust optimization based models require a set of potential realizations of uncertainty representation from a high-quality dataset of renewable energy resources [1], [2]. Thus, different methods have been proposed to generate adequate scenarios data for characterizing the distribution of renewable energy in optimization model and helping the power system more robust to the increasing uncertainty level, e.g., the stochastic economic dispatching [3] and the optimal decision-making of energy management [4], [5]. The scenarios data of renewable energy is used to build empirical probability distributions of density-based and metric-based ambiguity sets [6], [7], [8].

To generate scenarios data in a cheap and efficient way, the approaches can be categorized as the model-based and the model-free methods [9], [10]. The model-based methods are based on a set of model assumptions of probability distribution, and the distributions of scenarios data generated from them are estimated by specific sampling techniques, such as Monte Carlo sampling [11] and Latin Hypercube sampling [12]. Some of these methods further characterize the uncertainty via copula functions which can represent the correlation of multiple random variables and separate their strongly coupled joint distribution into multiple marginal distributions [13], [14]. Autoregressive-based models [15], [16], [17] are alternative approaches to model the time series with some assumptions for the distribution of forecasting error. They firstly identify the trend of historical time series, and then generates synthetic samples by additional random sampling.

On the contrary, the model-free methods do not require explicit model assumptions for the probability distribution of scenarios data. Compared to the model-based methods, they have stronger ability to fit data, but more computational cost and weaker explainability at the same time. Among them, Li et al. [18] proposed a hybrid deep-learning method, combining long short-term memory (LSTM) model, gated recurrent

TABLE I
SUMMARY OF RECENT STUDIES

References	Model-free	GAN-based	Data Missing	Multi-modal
[11]–[17]	-	-	-	-
[18], [19]	✓	-	-	-
[20]–[25]	✓	✓	-	-
[27], [28]	-	-	✓	-

unit (GRU) model and convolutional neural network (CNN), to generate weekly PV power scenarios data. Qi et al. [19] proposed an optimal configuration method for concentrating solar power based on variational autoencoder (VAE) model, which was used to generate day-ahead PV data. To generate scenarios data, Chen et al. [20], [21] used generative adversarial network (GAN) to learn the intrinsic probabilistic patterns via an unsupervised way. The learned patterns was further used to generate wind and solar time-series scenarios data. Based on the framework of GAN, Liang et al. [22] proposed a incremental way to train GAN model coupled with reinforcement learning. Qiao et al. [23] introduced an orthogonal regularization and a spectral normalization to stabilize the training process of Wasserstein GAN (WGAN) and improve the scenarios generation quality. Dong et al. [24] proposed a controllable GAN, which can generate scenarios data of solar energy via mutual information maximization, and even create a new generation pattern that are different from collected scenarios samples. Yuan et al. [25] proposed an effective style-based GAN to capture the complex diurnal pattern and seasonality difference of renewable power.

Moreover, the issue of data missing, resulting from various system failures and expansions of renewable energy power station [26], [27], was also considered in recent studies. Hu et al. [27] proposed a transfer component analysis method to approximate the scenarios data distribution of newly-built wind farms. They further proposed a joint distribution adaptation method to transfer the scenarios data information of existing data-rich wind farms to the cross-domain distribution of newly-built wind farms [28].

To clearly present the status of researches, the state-of-the-art works are summarized in TABLE I. In general, the definitions of scenarios generation for various application cases are remarkably different in the type of data, the shape of scenarios samples and data structure. To address this issue in a general way, the issue can be summarized as a two-sample problem, which is a test to determine if two groups of samples are from different probability distributions [29]. Thus, if given a piece of scenarios sample

$$X = \{x_{it} | i = 1, \ldots, N, t = 1, \ldots, T\}, \tag{1}$$

where N is the number of power units and T is the length of time-series trajectories. The goal of the scenarios generation is to capture the intrinsic features of the historic data, and then generate a set of potential realizations of future states, instead of just giving a forecasting [6], [20]. Thus, the proposed model should be able to generate scenarios, including both the observed ones and the potential ones. And the estimation distribution $P(\hat{X})$ and the observation distribution $P_{data}(X)$ both conform to the same distribution, namely $P(\hat{X}) = P_{data}(X)$, because they are generated from the same system. Thus, this can be a two-sample problem. Then, it brings two key questions, that is, how to measure the distance between $P(\hat{X})$ and $P_{data}(X)$, and how to generate scenarios sample $\hat{X}$.

For the first question, to avoid ambiguous definition and uncertain hyperparameter selection for various application scenarios, the GAN model is naturally considered as the basic framework to approximate the probability distribution of real scenarios data. The framework of GAN is used to solve the two-sample problem regarding the distributions $P(\hat{X})$ and $P_{data}(X)$.

For the second question, some aforementioned studies use the deep-learning models to generate scenarios data by neural network. However, the quality of the generated data is limited due to the absence of consideration to the multimodality of renewable energy resources, as shown in Table I. Actually, the trend of multimodality is inevitable and evolving with the development of multi-energy power systems and multi-type sensor technique [9], [30], [31]. Besides, these models are not robust to be applied for the missing data because the missing value cannot be recovered from other modal information [32].

Thus, the main contributions of this article are as follows:

1) A novel cross-modal method is proposed to generate scenarios data of renewable energy via a cross-modal Generative Adversarial Network (cGAN).
2) A cross-modal data fusion technique is proposed to fuse graph data generated by GPS and time-series power output data in renewable energy power system via a spatio-temporal transformer. The transformer uses these multi-modal data to generate scenarios data by an end-to-end way.
3) Extensive experiments are conducted. The results show that cGAN achieves the state-of-the-art (SOTA) scenarios generation performance, even though the observation data is missing at random. Furthermore, the results show that the spatio-temporal transformer can explain the contributions of different modal data for model performance improvement, and provide more training stability for cGAN. Finally, the results show that the generated scenarios data is equivalent to the real-world data in a case study of stochastic day-ahead economic dispatching on a wind power dataset.

The rest of this article is organized as follows. Section II presents all components of cGAN, and describes the detailed principles and practical implement procedure. Section III analyzes and discusses the experimental design and the results of comparative experiments. Section IV concludes this article.

II. METHODOLOGY

In this section, the scenarios generation issue is considered as a two-sample problem [29] and solved by the principle of cGAN, as shown in Fig. 1. And then, the principle of the cross-modal data fusion is introduced. Finally, the implementation and algorithm are presented.

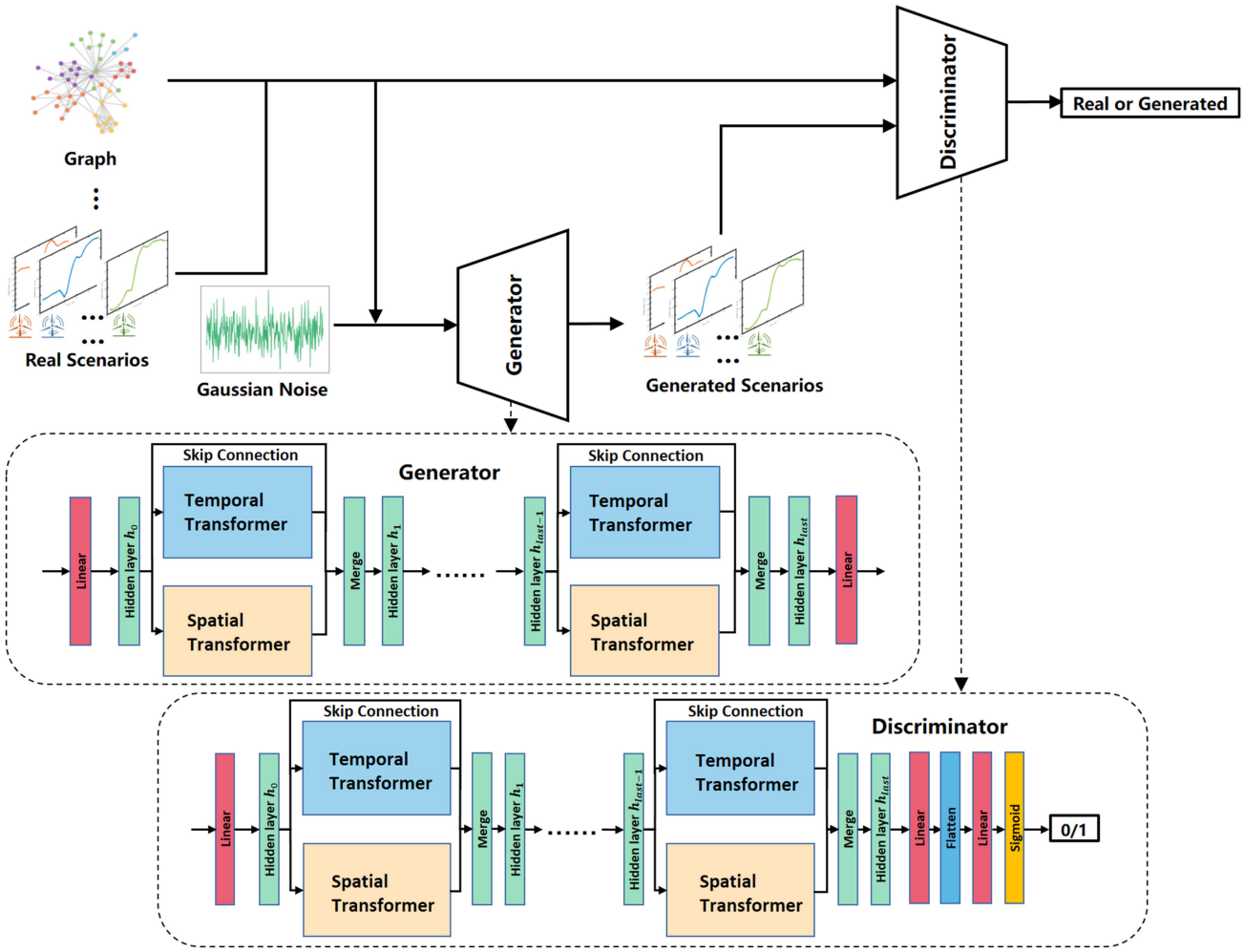


Fig. 1. Schematic of cGAN architecture.

A. Cross-Modal Generative Adversarial Network

To measure the discrepancy between $P(\hat{X})$ and $P_{data}(X)$, the first step is to relate the scenarios generation of renewable energy and the approximation principle of cGAN. The spatial relations of the power units are denoted as a graph $\mathcal{G}$ with an adjacency matrix A. By importing an extra control variable $Y \in \{0, 1\}$, all scenarios samples is relabeled to be observed samples while $Y = 1$, and generated scenarios samples while $Y = 0$. Then, the issue is simplified to find a distribution $Q(\hat{X}, \mathcal{G}, Y)$ to approximate $P(\hat{X}, \mathcal{G}, Y)$. The $Q(\hat{X}, \mathcal{G}, Y)$ is defined as

$$Q(\hat{X}, \mathcal{G}, Y) = \begin{cases} P_{data}(X, \mathcal{G})P_1, & Y = 1, \\ Q(\hat{X}, \mathcal{G})P_0, & Y = 0, \end{cases} \tag{2}$$

where P_1 and P_0 control the proportion of observed samples and generated samples. Without loss of generality, it is assumed that $P_1 = P_0 = \frac{1}{2}$. Then, the Kullback-Leibler divergence of $Q(\hat{X}, \mathcal{G}, Y)$ and $P(\hat{X}, \mathcal{G}, Y)$ is defined as

$$\begin{aligned} & D_{KL}(Q(\hat{X}, \mathcal{G}, Y)||P(\hat{X}, \mathcal{G}, Y)) \\ &= \int P_{data}(X, \mathcal{G})P_1 \log \frac{P_{data}(X, \mathcal{G})P_1}{P(Y = 1|\hat{X}, \mathcal{G})P_{data}(X, \mathcal{G})} d\hat{X} \\ &\quad + \int Q(\hat{X}, \mathcal{G})P_0 \log \frac{Q(\hat{X}, \mathcal{G})P_0}{P(Y = 0|\hat{X}, \mathcal{G})P_{data}(X, \mathcal{G})} d\hat{X} \\ &\propto \int P_{data}(X, \mathcal{G}) \log \frac{1}{P(Y = 1|\hat{X}, \mathcal{G})} d\hat{X} \\ &\quad + \int Q(\hat{X}, \mathcal{G}) \log \frac{Q(\hat{X}, \mathcal{G})}{P(Y = 0|\hat{X}, \mathcal{G})P_{data}(X, \mathcal{G})} d\hat{X}, \end{aligned} \tag{3}$$

where $P(Y = 0|\hat{X}, \mathcal{G}) = 1 - P(Y = 1|\hat{X}, \mathcal{G})$. Obviously, $D_{KL}(Q(\hat{X}, \mathcal{G}, Y)||P(\hat{X}, \mathcal{G}, Y)) = 0$ if and only if $Q(\hat{X}, \mathcal{G}, Y)$ is equal to $P(\hat{X}, \mathcal{G}, Y)$. Let $D(\hat{X}, \mathcal{G}) = P(Y = 1|\hat{X}, \mathcal{G})$, and $G(X, \mathcal{G}, Z) = Q(\hat{X}, \mathcal{G}) \sim P_G$, which is a cross-modal data generator. $Z \sim N(\mathbf{0}, \mathbf{I})$ is multivariate standard Gaussian noise sampling,

Then, according to [33], [34], the Eq. (3) can be naturally reduced to a min-max game, as follows:

$$\min_G \max_{D:||D||_L \leq 1} E_{X \sim P_{data}}[D(X, \mathcal{G})] - E_{\hat{X} \sim P_G}[D(\hat{X}, \mathcal{G})], \tag{4}$$

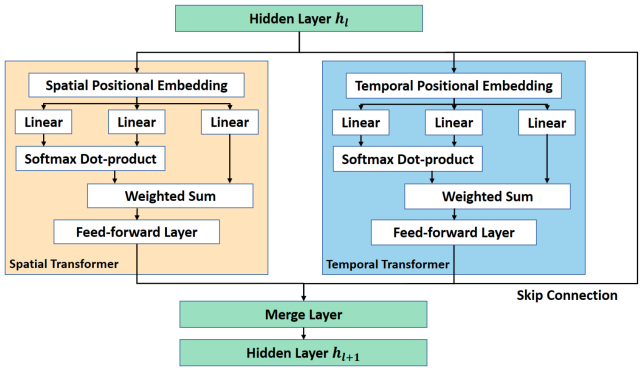


Fig. 2. $l+1$-th Spatio-temporal transformer module.

where $||D||_L \leq 1$ means 1-Lipschitz constraints for D. Then, according to [35], the loss functions of D and G can be rewritten as

$$L[D] = D(\hat{X}, \mathcal{G}) - D(X, \mathcal{G}) + \lambda(||\nabla_{\tilde{X}} D(\tilde{X}, \mathcal{G})|| - 1)^2,$$
$$L[G] = -D(\hat{X}, \mathcal{G}) + \mu L_R, \tag{5}$$

where $\hat{X} = G(X, \mathcal{G}, Z)$, and $\tilde{X} = \epsilon X + (1-\epsilon)\hat{X}$ in which ϵ conforms to uniform distribution $U(0,1)$. λ and μ are both free parameters. L_R is a representation error term defined as

$$L_R = \frac{1}{CI[\mathbf{I}_{it}=1]} \sum_{i,t} \mathbf{I}_{it} ||\hat{x}_{it} - x_{it}||^2, \tag{6}$$

where $\hat{x}_{it} \in \hat{X}$ is the generated value. $\mathbf{I}_{it} = 0$ if x_{it} is missing value in corrupted observation data, otherwise $\mathbf{I}_{it} = 1$. Moreover, $CI[\mathbf{I}_{it} = 1]$ is a countif function that counts the element number of $\mathbf{I}_{it} = 1$. When the game of D and G reaches Nash equilibrium through the alternating optimization process, P_G approaches to P_{data}, namely $P_{data}(X, \mathcal{G}) = Q(\hat{X}, \mathcal{G})$. Then, due to the fact that the graph $\mathcal{G}$ is static, the distribution of $\mathcal{G}$ is fixed and $Q(\mathcal{G}|\hat{X}) = P_{data}(\mathcal{G}|X)$. Finally, the two-sample problem is solved by $Q(\hat{X}) = \frac{Q(\hat{X},\mathcal{G})}{Q(\mathcal{G}|\hat{X})} = \frac{P_{data}(X,\mathcal{G})}{P_{data}(\mathcal{G}|X)} = P_{data}(X)$. The trained generator can be used to generate scenarios data of renewable energy from distribution P_{data}.

B. Cross-Modal Data Fusion

The cross-modal data fusion technique is proposed to fuse graph data generated by GPS and time-series power output data via the spatio-temporal transformer. The discriminator and generator both use the spatio-temporal transformer to parameterize as D_ω and G_θ, respectively. As shown in Fig. 1, the generator uses the graph $\mathcal{G}$, Gaussian noise sample Z and real scenarios X as input, and finally output the generated scenarios $\hat{X}$. The discriminator uses the graph $\mathcal{G}$ and the scenarios data (X or $\hat{X}$) as input, and finally output the value of label ("Real" or "Generated").

The spatio-temporal transformer is organized by stacked spatio-temporal transformer modules, as shown in Fig. 2. In the l-th transformer module, the initial input $X^{\mathcal{S}}$ of the spatial transformer is defined as

$$X^{\mathcal{S}} = h_l + E^{\mathcal{S}}, \tag{7}$$

where $h_l \in \mathbb{R}^{N\times T\times C}$ is the hidden feature achieved at the l-th transformer module and C is the number of expansion dimension. $E^{\mathcal{S}} = AW_a^{\mathcal{S}}$ is embedded firstly into $\mathbb{R}^{N\times T\times C}$, and then added to h_l. This process embeds the spatial modal information, namely the adjacency matrix A, into scaled positional code $E^{\mathcal{S}}$, and pass down to the final output layer-wisely. Then, according to the standard transformer model [36], the queries, keys and values are calculated by

$$Q^{\mathcal{S}} = X^{\mathcal{S}} W_q^{\mathcal{S}},$$
$$K^{\mathcal{S}} = X^{\mathcal{S}} W_k^{\mathcal{S}},$$
$$V^{\mathcal{S}} = X^{\mathcal{S}} W_v^{\mathcal{S}}, \tag{8}$$

where $W_q^{\mathcal{S}}, W_k^{\mathcal{S}}, W_v^{\mathcal{S}} \in \mathbb{R}^{C\times C}$ are corresponding matrices for $Q^{\mathcal{S}}, K^{\mathcal{S}}, V^{\mathcal{S}}$. Then, the dynamical spatial dependencies are captured by

$$M^{\mathcal{S}} = \text{softmax}\left(\frac{Q^{\mathcal{S}}(K^{\mathcal{S}})^T}{\sqrt{d_k^{\mathcal{S}}}}\right), \tag{9}$$

where softmax is a softmax dot-product function, and $d_k^{\mathcal{S}} = C$ is a scaling factor. $M^{\mathcal{S}}$ represents the extraction weights of each spatial features, and the spatial features are extracted by

$$U^{\mathcal{S}} = M^{\mathcal{S}} V^{\mathcal{S}}, \tag{10}$$

where $M^{\mathcal{S}}V^{\mathcal{S}}$ is matrix product at dimension N and T and $U^{\mathcal{S}} \in \mathbb{R}^{N\times T\times C}$. After that, a feed-forward layer is added as

$$Y^{\mathcal{S}} = \text{ReLU}(\text{ReLU}(U^{\mathcal{S}} W_0^{\mathcal{S}}) W_1^{\mathcal{S}}) W_2^{\mathcal{S}}, \tag{11}$$

where $W_0^{\mathcal{S}}, W_1^{\mathcal{S}}, W_2^{\mathcal{S}} \in \mathbb{R}^{C\times C}$ are the parameters of three feed-forward neural network layers, respectively.

Similarly, the temporal transformer also embeds the temporal modal information into positional code, and then add it to h_l. It can be defined as

$$X^{\mathcal{T}} = h_l + E^{\mathcal{T}}, \tag{12}$$

where $E^{\mathcal{T}} = \text{Embedding}([1, \ldots, T])$. Embedding$(\cdot)$ is a general embedding layer, which embeds $[1, \ldots, T]$ into a vector with fixed dimension. Then, the queries, keys, values of the temporal transformer are obtained by

$$Q^{\mathcal{T}} = X^{\mathcal{T}} W_q^{\mathcal{T}},$$
$$K^{\mathcal{T}} = X^{\mathcal{T}} W_k^{\mathcal{T}},$$
$$V^{\mathcal{T}} = X^{\mathcal{T}} W_v^{\mathcal{T}}, \tag{13}$$

where $W_q^{\mathcal{T}}, W_k^{\mathcal{T}}, W_v^{\mathcal{T}} \in \mathbb{R}^{C\times C}$ are corresponding matrices for $Q^{\mathcal{T}}, K^{\mathcal{T}}, V^{\mathcal{T}}$, respectively. And then,

$$M^{\mathcal{T}} = \text{softmax}\left(\frac{Q^{\mathcal{T}}(K^{\mathcal{T}})^T}{\sqrt{d_k^{\mathcal{T}}}}\right), \tag{14}$$

where $d_k^{\mathcal{T}} = C$, and $(Q^{\mathcal{T}}(K^{\mathcal{T}})^T$ is matrix product at dimension N and T, which means $M^{\mathcal{T}} \in \mathbb{R}^{N\times N\times C}$. Then,

$$U^{\mathcal{T}} = M^{\mathcal{T}} V^{\mathcal{T}}, \tag{15}$$

where $M^{\mathcal{T}}V^{\mathcal{T}}$ means matrix product at dimension N and T and $M^{\mathcal{T}} \in \mathbb{R}^{N\times T\times C}$. Then, a same feed-forward layer is applied as

$$Y^{\mathcal{T}} = \text{ReLU}(\text{ReLU}(U^{\mathcal{T}}W_0^{\mathcal{T}})W_1^{\mathcal{T}})W_2^{\mathcal{T}}, \tag{16}$$

where $W_0^{\mathcal{T}}, W_1^{\mathcal{T}}, W_2^{\mathcal{T}} \in \mathbb{R}^{C\times C}$ are the parameters of three feed-forward neural network layers, respectively.

Finally, a merge layer is added to join $Y^{\mathcal{S}}$, $Y^{\mathcal{T}}$ and h_l, as follow

$$h_{l+1} = \text{ReLU}([Y^{\mathcal{S}}; Y^{\mathcal{T}}; h_l]W_m), \tag{17}$$

where $[\cdot;\cdot;\cdot]$ means matrices concatenation at dimension C and $W_m \in \mathbb{R}^{3C\times C}$. Given the scenarios $X \in \mathbb{R}^{N\times T}$ and a graph $\mathcal{G}$ with the matrix $A \in \mathbb{R}^{N\times N}$, $\hat{X} \in \mathbb{R}^{N\times T}$ can be generated by the generator via random sampling. At the first, $Z \in \mathbb{R}^T$ is sampled by T-dimensional standard Gaussian distribution, and further expanded and added to the data X. Then, the input data X is scaled to be C-dimensional, namely as $\mathbb{R}^{N\times T\times C}$, by 1×1 convolution layer. The feature information flow goes through multiple transformer modules, and finally generates the scenarios $\hat{X}$. Then, the generated data and the observed data are both inputted to the discriminator with artificial labels $0, 1$. These data continue to go through multiple transformer modules, and finally output the predictive labels via

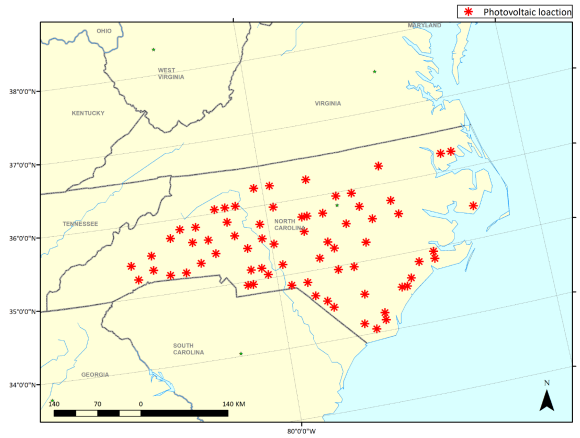
$$O_{pred} = \text{Sigmoid}(\text{Flatten}(h_{last}W_{d_0})W_{d_1}), \tag{18}$$

where h_{last} is the last hidden feature of these transformer modules, W_{d_0}, W_{d_1} are weighted matrices for two linear neural networks, and Flatten$(\cdot)$ is a function to flatten all values of dimension N, T and C.

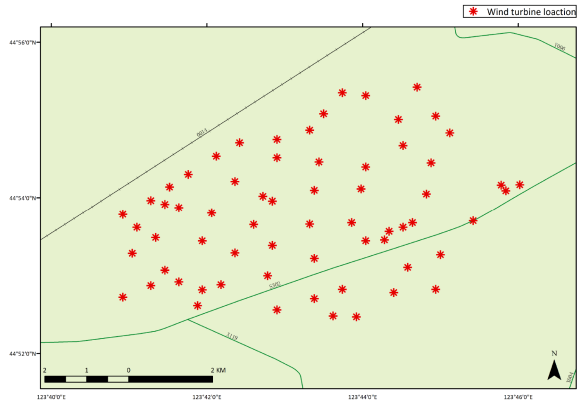
C. Implementation and Algorithm

The implementation of cGAN starts from parameter initialization of G_θ and D_ω by Adam optimizer [37] with learning rate α, as presented in Algorithm 1. After initialization, the algorithm begins selecting batch samples to train D_ω for n_{disc}-times iterations, and then train G_θ for n_{gen}-times iterations. In each iteration, random variable Z is resampled, and the graph $\mathcal{G}$ is invariant. After training G_θ, it repeats the above process n_{epoch} times, until the parameter of G_θ and D_ω are both converged. Finally, a trained scenarios generator G_θ is obtained. The G_θ can be used to generate new scenarios samples by repeated random sampling of Z.

In this process, the total time complexity of Algorithm 1 is $O((NT^2C^2 + N^2TC^2)(l_{gen} * n_{gen} + 2l_{disc} * n_{disc})n_{epoch})$, where l_{gen} and l_{disc} are assumed to be the number of spatio-temporal transformer modules for G_θ and D_ω, respectively. But in practice, due to the fact that $n_{epoch} \gg \max(N, T, C)$ and $\min(N, T, C) \gg 4 > \max(l_{gen}, l_{disc}, n_{gen}, n_{disc})$, the time complexity can be approximated as $\mathcal{O}((NT^2C^2 + N^2TC^2)n_{epoch})$. Thus, the computational cost of **Algorithm 1** is mainly up to the sample size N, T, C and the number of iterations n_{epoch}.



(a) The layout of PV power units, locating at range of $34.05°\text{N} \sim 36.25°\text{N}, 76.15°\text{W} \sim 82.65°\text{W}$



(b) The layout of Wind power units, locating at range of $44.875°\text{N} \sim 44.924°\text{N}, 123.682°\text{E} \sim 123.767°\text{E}$

Fig. 3. Location of renewable energy power sites.

III. Experiments

A. Description of Dataset

The performance of cGAN is verified on two public dataset, including a PV power dataset and a wind power dataset, respectively. Each dataset contains time-series observations of renewable energy and its corresponding geographic information of power units, as detailed bellow.

PV power dataset is a widely-used dataset [20], [21], [22], [23], [24], [25], collected from the publicly available solar power dataset from National Renewable Energy Laboratory (NREL) [38]. It contains the time-series PV power output data of the utility-scale PV sites, which is aggregated into 15-minute time granularity during whole 2006. All PV sites locate in North Carolina of the USA, and the spatial layout representing as longitude and latitude collected from GPS is shown in Fig. 3(a). The number of wind farm sites in the dataset is equal to 69, while the installed capacity ranges from 10 MW to 134 MW.

Wind power dataset is actually sampled at a wind farm in Jilin province of China. It contains 66 wind turbines (WTs), whose standard capacity are all 1500 kW, and each WT is considered as a power unit in this article. The time range of this dataset

Algorithm 1: Cross-Modal Adversarial Generation.

Input: Collected Data, parameterized G_θ and D_ω
Output: Trained scenarios generator G_θ

Initialize G_θ and D_ω by Adam optimizer with learning rate α;
while *not yet repeat* n_{epoch} *times* **do**
  for *Iteration = 1, 2, ...,* n_{disc} **do**
    Select a minibatch X from collected data;
    Sample $Z \sim N(\mathbf{0}, \mathbf{I})$, and $\epsilon \sim U(0,1)$;
    $\hat{X} \leftarrow G_\theta(X, \mathcal{G}, Z)$;
    $\tilde{X} \leftarrow \epsilon X + (1-\epsilon)\hat{X}$;
    $L[D_\omega] \leftarrow$
    $D_\omega(\hat{X}) - D_\omega(X) + \lambda(||\nabla_{\tilde{X}} D_\omega(\tilde{X})|| - 1)^2$;
    $\omega \leftarrow \text{Adam}(\nabla_\omega L[D_\omega])$;
  for *Iteration = 1, 2, ...,* n_{gen} **do**
    Select a minibatch X from collected data;
    Sample $Z \sim N(\mathbf{0}, \mathbf{I})$;
    $\hat{X} \leftarrow G_\theta(X, \mathcal{G}, Z)$;
    $L[G_\theta] \leftarrow -D_\omega(\hat{X}) + \mu L_R$,;
    $\theta \leftarrow \text{Adam}(\nabla_\theta L[G_\theta])$;
return G_θ

is 10 months, and all data is also aggregated into 15-minute time granularity. The GPS locations of these WTs are shown in Fig. 3(b).

B. Performance Evaluation

To evaluate the performance of baseline models, maximum mean discrepancy (MMD), widely used in [23], [27], [28], [39], [40], [41], is adopted to measure difference between the generated sample distribution and the expected one. In practice, we randomly split the whole dataset into training set and test set, and the test set reflects the fundamental property of global sample distribution. The goal of scenarios generation is to approximate the distribution of test set by remained training data. Thus the MMD is used to be the index defined as

$$MMD(\hat{X}, X_{te}) = \frac{1}{T(T-1)} \sum_{m \neq n}^{T} k(\hat{X}_{:m}, \hat{X}_{:n}) + k(X^{te}_{:m}, X^{te}_{:n}) - \frac{2}{T^2} \sum_{m=1}^{T} \sum_{n=1}^{T} k(\hat{X}_{:m}, X^{te}_{:n}), \quad (19)$$

where $k(x, y) = \exp(-\frac{1}{\gamma}||x-y||^2)$ is general Gaussian kernel function. $\hat{X}$ is the generated scenarios sample and X^{te} is the target test sample. The sample size of them is defined in (1).

To conduct general comparisons between different models on different datasets, a scaled MMD score is used to measure overall skill power. It is defined as

$$MMD = 1 - \frac{MMD(\hat{X}, X_{te}) - MMD(X_{tr}, X_{te})}{MMD(X_{tr}, X_{te})}, \quad (20)$$

where X^{tr} is the training samples. MMD score scales the MMD index to be in range of $0 \sim 1$ approximately, and treat each model and dataset fairly.

To calculate the contributions of spatial and temporal transformers based on MMD score, the performance improvements are defined as

$$d_S = MMD_{cGAN} - MMD_{cGAN-T},$$
$$d_T = MMD_{cGAN} - MMD_{cGAN-S}, \quad (21)$$

where cGAN-S and cGAN-T are the CGAN only with spatial transformer and only with temporal transformer, respectively. Then, the contributions are defined as

$$con_S = \frac{d_S}{d_S + d_T} \times 100\%,$$
$$con_T = \frac{d_T}{d_S + d_T} \times 100\%. \quad (22)$$

Furthermore, to test the stability of baseline models when collected data is corrupted, the training samples split from whole dataset are damaged randomly, namely as missing at random (MAR) [42]. The missing rate is generally defined as

$$\eta = \frac{CI[\mathbf{I}_{it} = 0]}{N \times T} \times 100\%, \quad (23)$$

where $\mathbf{I}_{it} = 0$ if x_{it} is missing value, otherwise $\mathbf{I}_{it} = 1$, and $CI[\mathbf{I}_{it} = 0]$ is the countif function that counts the number of missing value.

C. Model Settings

In the experiments, the generator and discriminator of cGAN are organized with two spatio-temporal transformer modules. The embedding channels C is uniformly set as 32, and the time-series length T is set as 96 for day-ahead scenarios samples with 15-minute time granularity. N is the number of power units, which is adapted to the selected dataset. The parameter γ is set as 1 in MMD formula (see (19)). Furthermore, the adjacency matrix A is obtained by

$$A_{ij} = \begin{cases} e^{-\frac{d_{ij}}{\sigma}} & , i \neq j, \\ 0 & , i = j, \end{cases} \quad (24)$$

where d_{ij} is geographic distance (km) between power units i, j, which is calculated by the data of longitude and latitude collected from GPS. σ is a scaling factor that scales d_{ij} into appropriate range. σ is set as 100 for PV power dataset, and set as 1 for wind power dataset.

To realize stable adversarial training, the iteration n_{disc} and n_{gen} in algorithm 1 is set as 3 and 2 at the PV power dataset, while 2 and 2 at the wind power dataset, respectively. The n_{epoch} is set as 1000. The learning rate $\alpha = 0.0001$, and batch size is set as 32. The free parameter λ and μ (see (5)) are both set as 10.

Moreover, according to the reviews, ARIMA [17], LSTM [18], GRU [18], VAE [19], WGAN [20] and C-StyleGAN2-SE [25] are selected to be the baseline models.

Among them, ARIMA is one of functional component used to generate base and spike component scenarios of electricity market in [17]. But in this article, it is adjusted to adapt to our issue and used to generate time-series scenarios data of renewable energy with additive Gaussian noise. The P, D, Q value, default parameter of ARIMA, are set as 4, 0, 0, respectively.

TABLE II
RESULTS OF MULTIPLE COMPARATIVE EXPERIMENTAL MODELS ON PV POWER DATASET

Missing Rate/MMD/Models	ARIMA	LSTM	GRU	VAE	WGAN	C-StyleGAN2-SE	cGAN-S	cGAN-T	cGAN
0%	0.519	0.881	0.890	0.727	0.786	0.899	0.892 (con_S : 87.9%)	0.823 (con_T : 12.1%)	**0.903**
5%	0.511	0.877	0.883	0.692	0.782	0.893	0.890 (con_S : 95.3%)	0.813 (con_T : 4.7%)	**0.894**
10%	0.510	0.872	0.875	0.691	0.781	0.888	0.886 (con_S : 93.2%)	0.810 (con_T : 6.8%)	**0.892**
15%	0.505	0.871	0.869	0.682	0.775	0.870	0.883 (con_S : 95.1%)	0.809 (con_T : 4.9%)	**0.887**
20%	0.504	0.866	0.868	0.673	0.772	0.868	0.879 (con_S : 97.4%)	0.805 (con_T : 2.6%)	**0.881**

The bold entities indicate the performance or improvement of our method.

TABLE III
RESULTS OF MULTIPLE COMPARATIVE EXPERIMENTAL MODELS ON WIND POWER DATASET

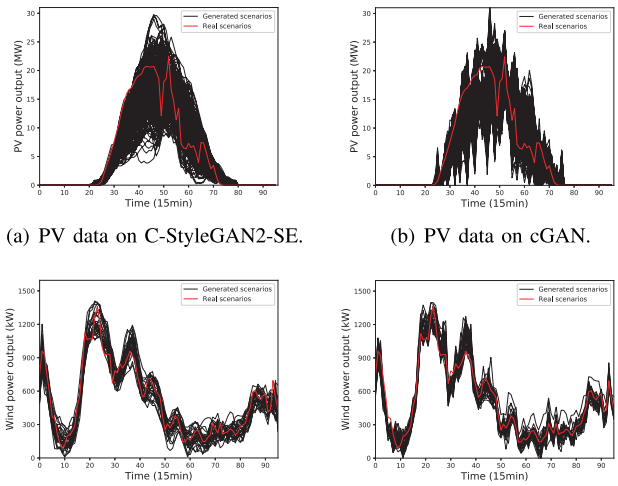
Missing Rate/MMD/Models	ARIMA	LSTM	GRU	VAE	WGAN	C-StyleGAN2-SE	cGAN-S	cGAN-T	cGAN
0%	0.621	0.919	0.932	0.784	0.839	0.929	0.933 (con_S : 93.5%)	0.852 (con_T : 6.5%)	**0.939**
5%	0.620	0.898	0.914	0.701	0.798	0.914	0.930 (con_S : 98.8%)	0.849 (con_T : 1.2%)	**0.931**
10%	0.617	0.851	0.895	0.682	0.782	0.903	0.925 (con_S : 95.7%)	0.840 (con_T : 4.3%)	**0.929**
15%	0.615	0.849	0.882	0.689	0.751	0.881	0.909 (con_S : 91.9%)	0.837 (con_T : 8.1%)	**0.916**
20%	0.612	0.824	0.859	0.677	0.742	0.874	0.890 (con_S : 98.4%)	0.831 (con_T : 1.6%)	**0.891**

The bold entities indicate the performance or improvement of our method.

LSTM and GRU has 2 hidden layers for bidirectional time-series feature extraction, and sequentially connect a linear layer with N neural units for fitness of extracted features and output. The trained LSTM and GRU model can generate scenarios samples by inputting real samples, which is added with standard Gaussian noise term.

VAE is organized with encoder and decoder, and they are both parameterized by fully-connected neural network, which has 3 hidden layers with neural units [2048, 2048, 2048] and leaky ReLU activation with 0.2 slope. VAE model firstly encodes input samples to be two 384-dimensional vectors, which represents mean and variance of Gaussian latent variables. And then, it resamples these variables to generate new scenarios samples via decoder.

In WGAN, the model setting is almost same as our proposed cGAN except for the spatio-temporal transformer structure and the representation error term. The generator and discriminator are parameterized by fully-connected neural networks, which are same as that of VAE. The only difference is that the last hidden layer of discriminator is connected with an output layer to map samples to single real value. C-StyleGAN2-SE is set with the same parameters as the relevant article. Note that, the other unmentioned parameters of all baseline models are also same as the relevant papers if there is no special statement.



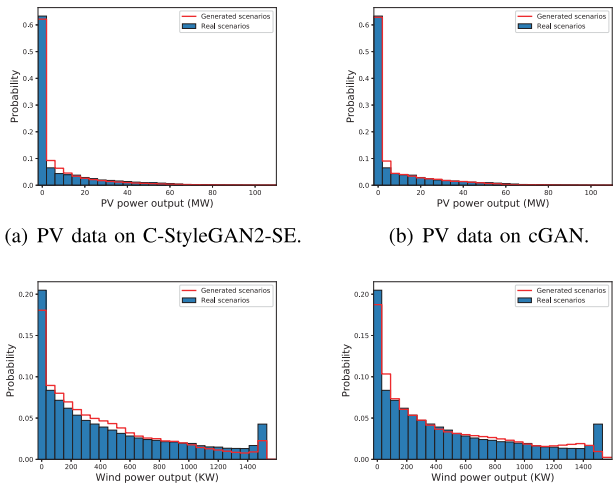
(a) PV data on C-StyleGAN2-SE. (b) PV data on cGAN.

(c) WT data on C-StyleGAN2-SE. (d) WT data on cGAN.

Fig. 4. Comparisons on the day-ahead scenarios data generated by cGAN and C-StyleGAN2-SE.

D. Scenarios Generation of Renewable Energy

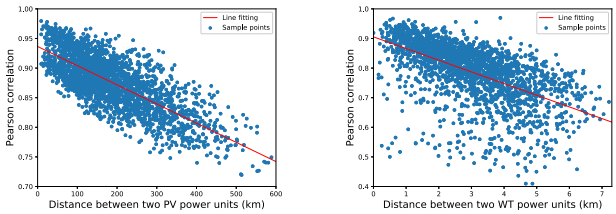
Extensive experiments are conducted to verify the performance of cGAN, compared with other baseline models from a comprehensive perspective. As presented in Tables II and III, the MMD score of cGAN achieved at two datasets are both higher than other models at data missing rate 0%. Moreover, with the increasing missing rate, the MMD scores of all models decline at different degrees. By contrast, the MMD score of cGAN still keeps outstanding, and especially outperforms the second best model, namely as C-StyleGAN2-SE, with 1.5% improvement at PV power dataset and 1.9% improvement at wind power dataset when missing rate is 20%. Note that, the performances of these baseline models decline faster than cGAN as data missing rate increases. Thus, cGAN achieves the SOTA performance, and has better robustness for data missing than other baseline models.



(a) PV data on C-StyleGAN2-SE. (b) PV data on cGAN.

(c) WT data on C-StyleGAN2-SE. (d) WT data on cGAN.

Fig. 5. Comparisons on the distribution of scenarios data generated by cGAN and C-StyleGAN2-SE.



(a) PV power dataset. (b) Wind power dataset.

Fig. 6. Time-series Pearson correlation of two power units is negatively related to the distance of them.

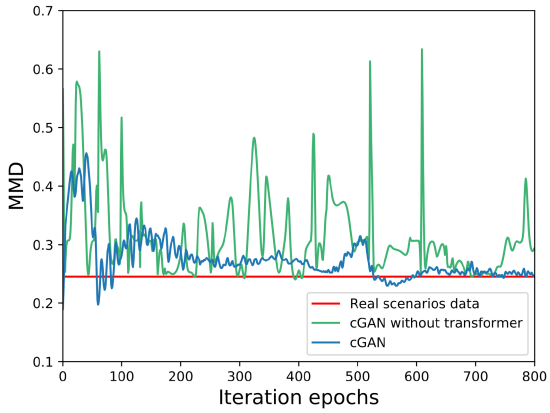


Fig. 7. MMD index convergence of our model with and without spatio-temporal transformer.

Then, to further present the advantages of cGAN for realistic scenarios generation, the comparisons on the day-ahead scenarios data generated by cGAN and C-StyleGAN2-SE are shown in Fig. 4. By visual inspection, the generated PV and Wind power energy scenarios on the two models both mostly

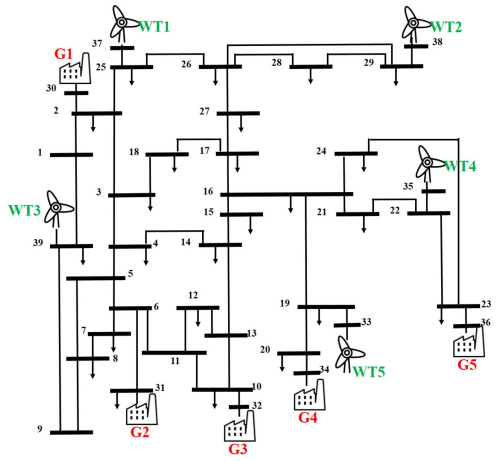


Fig. 8. Modified IEEE-39 bus test system.

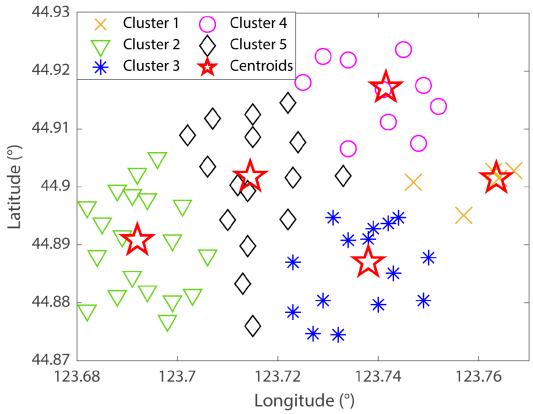


Fig. 9. Results of WT location clustering.

TABLE IV
CLUSTERS OF THE 66 WTS

Cluster	Bus Loc.	# WT
1	37	1-13, 18, 19, 20, 21, 26, 57
2	38	37, 40, 43-45, 47-52, 54, 55, 66
3	39	17, 24, 25, 27, 29-31, 34, 35, 38, 39, 46, 53
4	35	22, 28, 32, 33, 36, 41, 42, 61, 63-65
5	33	14-16, 23, 56, 58-60, 62

cover the real scenarios data. It shows the feasibility to use the generated scenarios to infer the correct day-ahead renewable energy patterns. However, the generated scenarios of cGAN has narrower diversity compared with that of the C-StyleGAN2-SE model, especially on the Wind power dataset. That is, the real scenarios can be searched from a narrower range of potential scenarios realization.

Moreover, the results of probabilistic approximation are shown in Fig. 5. It is clear that the distribution of scenarios data generated from C-StyleGAN2-SE deviates from that of the real scenarios data, especially for lower power output value. That is, C-StyleGAN2-SE is not effective for the scenarios

TABLE V
RESULTS OF STOCHASTIC DAY-AHEAD ECONOMIC DISPATCHING

	Raw	ARIMA	LSTM	GRU	VAE	WGAN	C-StyleGAN2-SE	Our Model
Expected Cost ($)	93,626,436	86,549,168 (**-7.56%**)	90,088,410 (**-3.78%**)	91,809,385 (**-1.94%**)	87,192,371 (**-6.87%**)	88,200,425 (**-5.80%**)	91,897,417 (**-1.85%**)	92,291,324 (**-1.43%**)

The bold entities indicate the performance or improvement of our method.

generation of low power output, and even no power output. By contrast, the distribution of scenarios data generated from cGAN almost coincides with that of the real scenarios data, which means $P(\hat{X})$ is close to $P_{data}(X)$. Thus, cGAN has comprehensive advantages compared with the other baseline models.

E. Effectiveness of Cross Modal

To test the effectiveness of cross modal, the relations between time-series Pearson correlation and spatial distance are presented in the Fig. 6. The figure shows that the time-series Pearson correlation of two power units is negatively related to the distance of them whatever on PV power dataset and wind power dataset. That is, if the value of time series is missing, the missing value of the power unit can be recovered by the time series of other power units with high correlation to this unit, while the correlation can be measured by the distance of them. Thus, the cross modal is essential for scenarios generation due to the fact that the spatial and temporal modal data can provide more supplementary information for the missing value, respectively.

To quantify their contributions, one experiment is conducted to test the MMD score of cGAN only with spatial transformer or only with temporal transformer, recorded as cGAN-S and cGAN-T. The results in Tables II and III show that the spatial transformer provides much larger contributions to the performance improvement of cGAN than the temporal transformer. The conclusion can be generalized to the spatial and temporal modal.

Furthermore, one ablation experiment is conducted to compare the training efficiency of cGAN with and without the transformer. For cGAN without transformer module, the generator and discriminator of cGAN are both parameterized by 3-layers fully-connected neural networks. Before implementing this experiment, data samples is randomly divided into training and test set, which is considered as real scenarios data. The MMD index of them, namely $MMD(X_{tr}, X_{te})$, is calculated to be a certain value. Without loss of generality, its value is 0.2456 for general validation. As shown in Fig. 7, the MMD index of generated data and test set, namely $MMD(\hat{X}, X_{te})$, is calculated after each iteration epoch of the two training processes. Clearly, the training process of cGAN is more stable than that of cGAN without transformer. At the later training process, the value of MMD index is gradually converge to the red line. But in the training process of cGAN without the transformer, the value of MMD index fluctuates dramatically, and only stabilizes near the red line for few times. Thus, the spatio-temporal transformer of cGAN contributes to provide more training stability.

F. Scenarios Generation of Stochastic Day-Ahead Economic Dispatching

To further test the feasibility in practice, an experiment of stochastic day-ahead economic dispatching [43] is conducted on the modified IEEE 39-bus system to further validate the effectiveness of cGAN in real application scenarios. The test system contains five WTs which are aggregated from 66 WTs by their spatial locations, as shown in Fig. 8. The distribution of cluster centers is shown in Fig. 9, and all WTs are allocated to one of clusters uniformly as shown in Table IV. The dispatching problem is addressed by using sample average approximation method [44]. The objective function to minimize expected cost is defined as follows:

$$\min \sum_{t\in\mathbf{T}} \left(\sum_{i\in\mathbf{CG}} C^{uc}_{i,t} + \frac{1}{N_s} \sum_{s\in\mathbf{S}} \sum_{i\in\mathbf{CG}} C^{ed}_{i,s,t} \right) \tag{25}$$

In this equation, the first term represents the total unit commitment cost, and the second term represents the expected cost of the operating cost of the coal-fired generators using the sample average approximation method [44]. The settings about each variable can be seen in the supplement materials [45].

In the experiment, 5000 wind power scenarios samples within the range of 30 days are aggregated by different methods, respectively. Then, these scenarios samples and raw scenarios samples are all applied to the dispatching process, and the total expected costs of all methods are shown in Table V. It is clear that the expected cost of scenarios samples generated by cGAN is closest to that of raw scenarios samples with only 1.43% error. Consequently, the scenarios samples of cGAN achieve the highest generation quality.

IV. CONCLUSION

To address the issue of renewable energy scenarios generation, a novel cross-modal method is proposed to generate high-quality scenarios data from graph data generated by GPS and time-series power output data via cGAN. The proposed cGAN framework uses cross-modal data fusion technique to fuse the spatio-temporal modal data via the proposed spatio-temporal transformer, and finally output scenarios data in an end-to-end way.

To test the performance of cGAN, extensive experiments are conducted on a PV power dataset and a wind power dataset, compared with other baseline models, including ARIMA, LSTM, GRU, VAE, WGAN, C-StyleGAN2-SE, and so on. The experimental results show that cGAN achieves the SOTA performance

for scenarios generation, and even though the data is missing at random. Furthermore, one experiment is conducted to quantify the contributions between the different modals. The results show that the spatial transformer provides more performance improvement for cGAN than the temporal transformer. This conclusion can be generalized to the spatial and temporal modal data. To further present the importance of the spatio-temporal transformer, the ablation experiment is conducted to test the performance of cGAN with and without the transformer module. The results show that the spatio-temporal transformer provides more training stability for cGAN. Finally, the experiment of stochastic day-ahead economic dispatching is conducted for the verification of application value. The results show that the expected cost of scenarios samples generated by cGAN is closest to that of raw scenarios samples with only 1.43% error. That is, the generation samples is qualified enough to be applied for practical application. In order to improve the data quality of scenarios generation, future work will focus on the cross-modal data generation of renewable energy with richer multi-modal data representations and more complex running patterns.

REFERENCES

[1] L. Liu, Z. Hu, X. Duan, and N. Pathak, "Data-driven distributionally robust optimization for real-time economic dispatch considering secondary frequency regulation cost," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4172–4184, Sep. 2021.

[2] K. Nam, S. Hwangbo, and C. Yoo, "A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A. case study of korea," *Renewable Sustain. Energy Rev.*, vol. 122, 2020, Art. no. 109725.

[3] F. Lezama, J. Soares, P. Hernandez-Leal, M. Kaisers, T. Pinto, and Z. Vale, "Local energy markets: Paving the path toward fully transactive energy systems," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 4081–4088, Sep. 2019.

[4] Y. Li et al., "Optimal stochastic operation of integrated low-carbon electric power, natural gas, and heat delivery system," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 273–283, Jan. 2018.

[5] B. Odetayo, M. Kazemi, J. MacCormack, W. D. Rosehart, H. Zareipour, and A. R. Seifi, "A chance constrained programming approach to the integrated planning of electric power generation, natural gas network and storage," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6883–6893, Nov. 2018.

[6] B. Rachunok, A. Staid, J.-P. Watson, and D. L. Woodruff, "Assessment of wind power scenario creation methods for stochastic power systems operations," *Appl. Energy*, vol. 268, 2020, Art. no. 114986.

[7] G. Pan, W. Gu, Y. Lu, H. Qiu, S. Lu, and S. Yao, "Optimal planning for electricity-hydrogen integrated energy system considering power to hydrogen and heat and seasonal storage," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2662–2676, Oct. 2020.

[8] X. Zheng and H. Chen, "Data-driven distributionally robust unit commitment with Wasserstein metric: Tractable formulation and efficient solution method," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4940–4943, Nov. 2020.

[9] J. Li, J. Zhou, and B. Chen, "Review of wind power scenario generation methods for optimal operation of renewable energy systems," *Appl. Energy*, vol. 280, 2020, Art. no. 115992.

[10] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A. review," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 209–220, 2020.

[11] N. Amjady, J. Aghaei, and H. A. Shayanfar, "Stochastic multiobjective market clearing of joint energy and reserves auctions ensuring power system security," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1841–1854, Nov. 2009.

[12] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1319–1327, Aug. 2008.

[13] M. Sun, J. Cremer, and G. Strbac, "A novel data-driven scenario generation framework for transmission expansion planning with high renewable energy penetration," *Appl. Energy*, vol. 228, pp. 546–555, 2018.

[14] S. Camal, F. Teng, A. Michiorri, G. Kariniotakis, and L. Badesa, "Scenario generation of aggregated wind, photovoltaics and small hydro production for power systems applications," *Appl. Energy*, vol. 242, pp. 1396–1406, 2019.

[15] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Appl. Energy*, vol. 87, no. 3, pp. 843–855, 2010.

[16] Q. Tu et al., "Forecasting scenario generation for multiple wind farms considering time-series characteristics and spatial-temporal correlation," *J. Modern Power Syst. Clean Energy*, vol. 9, no. 4, pp. 837–848, 2021.

[17] D. Xiao and W. Qiao, "Hybrid scenario generation method for stochastic virtual bidding in electricity market," *CSEE J. Power Energy Syst.*, vol. 7, no. 6, pp. 1312–1321, 2021.

[18] H. Li, Z. Ren, Y. Xu, W. Li, and B. Hu, "A multi-data driven hybrid learning method for weekly photovoltaic power scenario forecast," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 91–100, Jan. 2022.

[19] Y. Qi, W. Hu, Y. Dong, Y. Fan, L. Dong, and M. Xiao, "Optimal configuration of concentrating solar power in multienergy power systems with an improved variational autoencoder," *Appl. Energy*, vol. 274, 2020, Art. no. 115124.

[20] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3265–3275, May 2018.

[21] Y. Chen, X. Wang, and B. Zhang, "An unsupervised deep learning approach for scenario forecasts," in *Proc. IEEE Power Syst. Computat. Conf.*, 2018, pp. 1–7.

[22] J. Liang and W. Tang, "Sequence generative adversarial networks for wind power scenario generation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 110–118, Jan. 2020.

[23] J. Qiao, T. Pu, and X. Wang, "Renewable scenario generation using controllable generative adversarial networks with transparent latent space," *CSEE J. Power Energy Syst.*, vol. 7, no. 1, pp. 66–77, 2020.

[24] W. Dong, X. Chen, and Q. Yang, "Data-driven scenario generation of renewable energy production based on controllable generative adversarial networks with interpretability," *Appl. Energy*, vol. 308, Art. no. 118387, 2022.

[25] R. Yuan, B. Wang, Y. Sun, X. Song, and J. Watada, "Conditional style-based generative adversarial networks for renewable scenario generation," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1281–1296, Mar. 2023.

[26] M. H. Kapourchali, M. Sepehry, and V. Aravinthan, "Fault detector and switch placement in cyber-enabled power distribution network," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 980–992, Mar. 2018.

[27] J. Hu, H. Li, and Z. Liu, "A novel scenario generation framework based on the knowledge of existing wind power plants," *IEEE Trans. Sustain. Energy*, vol. 12, no. 2, pp. 1229–1241, Apr. 2021.

[28] J. Hu and H. Li, "A transfer learning-based scenario generation method for stochastic optimal scheduling of microgrid with newly-built wind farm," *Renewable Energy*, vol. 185, pp. 1139–1151, 2022.

[29] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A Kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, 2012, Art. no. 723773.

[30] X. Fu, Q. Guo, H. Sun, Z. Pan, W. Xiong, and L. Wang, "Typical scenario set generation algorithm for an integrated energy system based on the wasserstein distance metric," *Energy*, vol. 135, pp. 153–170, 2017.

[31] G. Bedi, G. K. Venayagamoorthy, R. Singh, R. R. Brooks, and K.-C. Wang, "Review of internet of things (IoT) in electric power and energy systems," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 847–870, Apr. 2018.

[32] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019.

[33] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 2672–2680.

[34] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[35] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 5769–5779.

[36] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–13.
[38] NREL, "Solar power data for integration studies," Nov. 2022. [Online]. Available: https://www.nrel.gov/grid/solar-power-data.html#panel2
[39] F. Jamil, T. Verstraeten, A. Nowé, C. Peeters, and J. Helsen, "A deep boosted transfer learning method for wind turbine gearbox fault detection," *Renewable Energy*, vol. 197, pp. 331–341, 2022.
[40] I. Oyewole, A. Chehade, and Y. Kim, "A controllable deep transfer learning network with multiple domain adaptation for battery state-of-charge estimation," *Appl. Energy*, vol. 312, 2022, Art. no. 118726.
[41] J. Lin, J. Ma, J. Zhu, and H. Liang, "Deep domain adaptation for non-intrusive load monitoring based on a knowledge transfer learning network," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 280–292, Jan. 2022.
[42] H. Demirhan and Z. Renwick, "Missing value imputation for short to mid-term horizontal solar irradiance data," *Appl. Energy*, vol. 225, pp. 998–1012, 2018.
[43] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2197–2206, Nov. 2011.
[44] W. Tang and Y. J. Zhang, "A model predictive control approach for low-complexity electric vehicle charging scheduling: Optimality and scalability," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1050–1063, Mar. 2017.
[45] M. Kang, "Supplmentary materials," Nov. 2022. [Online]. Available: https://drive.google.com/file/d/1h7qt9I5r3-UylH7c-qwLL1MZW_D6qxos/view?usp=share_link

Mingyu Kang received the master's degree in computer technology in 2021 from Southeast University, Nanjing, China, where he is currently working toward the Ph.D. degree with the Jiangsu Key Laboratory of Networked Collective Intelligence, School of Cyber Science and Engineering. His research interests include complex system, complex network, deep learning, and causal inference.

Ran Zhu received the B.S. and M.S. degrees in electrical engineering from Nanjing Normal University, Nanjing, China, in 2012 and 2015, respectively. He is currently working toward the Ph.D. degree in cyber science and engineering in Southeast University, Nanjing. Before that, he was an Electrical Engineer with State Grid Electric Power Research Institute (SGEPRI), Wuhan, China. His research interests include modeling, planning, and optimization of resilient power system.

Duxin Chen (Member, IEEE) received the B.S. degree in automatic control and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2013 and 2018, respectively. He is currently an Associate Professor with the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, School of Mathematics, Southeast University, Nanjing, China. His research interests include complex networks and systems, artificial intelligence related theory and applications.

Chaojie Li (Member, IEEE) received the B.Eng. degree in electronic science and technology and the M.Eng. degree in computer science from Chongqing University, Chongqing, China, in 2007 and 2011, respectively, and the Ph.D. degree from RMIT University, Mel-bourne, VIC, Australia, in 2017, where he was a Research Fellow for one and a half years. He was a Senior Algorithm Engineer with Alibaba Group. His research interests include graph representation learning, distributed optimization and control in smart grid, neural networks, and their applications. He is a Senior Research Associate with thet School of Electrical Engineering and Telecommunications, UNSW. He was the recipient of ARC Discovery Early Career Researcher Award in 2020.

Wei Gu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2001 and 2006, respectively. From 2009 to 2010, he was a Visiting Scholar with the Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA. He is currently a Professor with the School of Electrical Engineering, Southeast University. He is the Director of the Institute of distributed generations and active distribution networks. His research interests include distributed generations and microgrids, integrated energy systems. He is the Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, *IET Energy Systems Integration* and *Automation of Electric Power Systems* (China).

Xusheng Qian received the graduation degree from the Harbin Institute of Technology, Harbin, China, in 1998. He is currently a Senior Engineer with the State Grid Jiangsu Electric Power Company Limited Marketing Service Center. His research interests include electric power marketing inspection and power consumption anomaly detection.

Wenwu Yu (Senior Member, IEEE) received the B.Sc. degree in information and computing science and the M.Sc. degree in applied mathematics from the Department of Mathematics, Southeast University, Nanjing, China, in 2004 and 2007, respectively, and the Ph.D. degree from the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China, in 2010. He is currently the Founding Director of the Laboratory of Cooperative Control of Complex Systems and the Deputy Associate Director of Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, an Associate Dean with the School of Mathematics, and a Full Professor with the Endowed Chair Honor with Southeast University. He held several visiting positions in Australia, China, Germany, Italy, The Netherlands, and the USA. He publishes about 100 IEEE TRANSACTIONS journal papers with more than 20000 citations. His research interests include multiagent systems, complex networks and systems, disturbance control, distributed optimization, machine learning, game theory, cyberspace security, smart grids, intelligent transportation systems, and big-data analysis. Dr. Yu was the recipient of the Second Prize of State Natural Science Award of China in 2016. He is also the Cheung Kong Scholars Programmer of Ministry of Education of China (Artificial Intelligence). He was listed by Clarivate Analytics/Thomson Reuters Highly Cited Researchers in Engineering during 2014–2022. He is an Editorial Board Member of several flag journals, including IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-PART II: EXPRESS BRIEFS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL CYBER-PHYSICAL SYSTEMS, *Science China Information Sciences*, and *Science China Technological Sciences*.