Enhancing Multimodal Meteorological Data Resolution via Diffusion Model for Accurate PV Potential Estimation

Jiaze Li[1], Zhiling Guo[2*], Huan Zhao[2], Hongjun Tan[2], Qing Yu[3], Rui Zhang[1], Jian Xu[2], Jinyue Yan[2]

1 Beijing-Dublin International College, Beijing University of Technology, Beijing, China

2 Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

3 School of Urban Planning & Design, Peking University

(Corresponding Author: zhiling.guo@polyu.edu.hk)

ABSTRACT

High-resolution meteorological data is crucial for accurately assessing photovoltaic potential at the microclimatic level. However, due to the scarcity of high-resolution data, modeling microclimates often becomes severely inaccurate, highlighting the need for technologies that improve the resolution of solar data. Additionally, current super-resolution models often fall short in accurately processing multimodal solar energy data, which is essential for improving the precision of PV potential estimation. This paper proposes a novel downscaling framework for meteorological data that leverages deep generative models. By integrating multimodal meteorological data from the National Solar Radiation Database (NSRDB), including Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI) and other relevant variables, with a diffusion model, the framework produces meteorological data at a high spatial resolution. Experimental results demonstrate that our super-resolution approach not only outperforms baseline methods but also significantly enhances the accuracy of PV potential estimation when compared to using coarse resolution data. Consequently, the diffusion-based super-resolution framework shows great promise for widespread adoption in the field of photovoltaic energy.

Keywords: Renewable energy, super resolution, diffusion model, multimodal learning, high resolution meteorological data, PV potential estimation

NONMENCLATURE

Abbreviations	
PV	Photovoltaic
DHI	Diffused Horizontal Irradiance
GHI	Global Horizontal Irradiance
DNI	Direct Normal Irradiance

1. INTRODUCTION

Photovoltaic (PV) power generation is a cornerstone of the global transition to renewable energy, offering a clean and sustainable alternative to fossil fuels. Over the past few decades, advancements in PV technology have significantly increased the efficiency and scalability of solar power systems, making them a vital component in the quest to reduce greenhouse gas emissions and combat climate change. The continued development of PV systems has enabled their widespread adoption across diverse geographic regions, contributing to the growth of renewable energy capacity worldwide[1].

As PV systems become increasingly integral to global energy strategies, accurately estimating their potential in specific locations is essential. Precise assessments of PV potential inform crucial decisions regarding site selection, system design, and energy policy, ultimately influencing the effectiveness and reliability of solar energy deployment. Accurate estimation depends on high-resolution solar data and the integration of multimodal inputs, such as temperature, wind speed, and solar irradiance.

However, current research efforts often encounter significant challenges in effectively utilizing this high-resolution, multimodal solar data. These challenges can lead to imprecise estimates of PV system performance, undermining the reliability of renewable energy projections and hindering the optimization of solar power generation.

Recent advancements in deep learning, particularly diffusion modeling, offer promising solutions to these challenges. Diffusion models, originally developed for generative tasks, have emerged as powerful tools in enhancing data resolution through a process of iteratively refining data representations. In the context

---

This is a paper for the 16th International Conference on Applied Energy (ICAE2024), Sep. 1-5, 2024, Niigata, Japan.

of super-resolution, diffusion models transform low-resolution data into high-resolution outputs by gradually denoising and upsampling the data through a series of learned transformations.

These models have shown significant potential in enhancing the resolution of various types of weather data, including solar irradiance, temperature, and wind speed. By refining the resolution of this data, diffusion models enable more precise detection and estimation of photovoltaic potential, which is crucial for optimizing the performance of PV systems.

This research specifically explores the use of diffusion modeling to achieve super-resolution of weather data, utilizing datasets from the National Solar Radiation Database (NSRDB). A key innovation in this work is the introduction of the DySample module within the upsampling process of the Unet architecture, coupled with the integration of the Swin Transformer technique to enhance feature representation during the denoising phase of the diffusion model. These advancements aim to improve the accuracy of model inference, thereby enhancing the precision of photovoltaic power generation estimates.

In summary, our contributions are:

a) Proposes a framework that improves PV potential estimation accuracy by fusing multimodal meteorological data with super-resolution techniques.

b) Utilizes an advanced diffusion model to achieve super-resolution for meteorological data, enhancing the resolution and detail.

c) Combines the Swin Transformer with the DySample module in the Unet's upsampling process during diffusion model denoising, leading to more precise inference and improved model performance.

2. RELATED WORK

2.1 PV potential estimation

Photovoltaic (PV) power generation is a critical element in the global transition to sustainable energy, offering a clean and renewable alternative to fossil fuels. As PV systems become more widespread, their potential to meet significant portions of global energy demand has grown, underscoring the importance of accurate planning and optimization. Central to this process is the precise estimation of PV potential, which involves assessing the capacity of a specific location to generate solar power based on various environmental and geographic factors. This estimation is crucial for determining the viability and efficiency of PV installations, guiding decisions related to site selection, system design, and energy policy[1].

Traditional methods for PV potential estimation have heavily relied on remote sensing (RS) technologies, which provide large-scale observational data across diverse environments. RS techniques, including satellite imagery, aerial photography, and LiDAR, have been widely used to assess solar energy potential due to their ability to capture extensive data over large areas. These methods, as highlighted by Chen et al[2], play a significant role in various stages of PV system development, from site selection to performance monitoring. However, the reliance on low-resolution data often limits these methods' ability to capture the microclimatic nuances that are essential for accurate PV potential assessments. This limitation can lead to inaccuracies in predicting PV system performance, thereby affecting the optimization and deployment of solar energy resources[3].

To address these challenges, there is a growing need for methods that enhance the resolution of data used in PV potential estimation. Improving data resolution allows for more detailed and accurate assessments, which are critical for optimizing PV system performance and maximizing the efficiency of solar energy generation.

2.2 Super resolution for meteorological data

In recent years, deep learning has emerged as a powerful tool for enhancing the resolution of image data, which is critical in fields such as photovoltaic (PV) potential estimation. Traditional super-resolution methods have primarily relied on techniques like Generative Adversarial Network (GAN) [4]. GANs work by pitting two neural networks against each other: a generator, which attempts to create high-resolution images from low-resolution inputs, and a discriminator, which evaluates the authenticity of these images. While GANs have shown some success in super-resolution tasks, they come with notable limitations, particularly in terms of achieving improved accuracy in PV potential estimation through super-resolution of solar-related data. GANs are often difficult to train, requiring extensive computational resources and careful tuning to avoid issues such as mode collapse [5].

To address these challenges, diffusion models have emerged as a promising alternative for super-resolution tasks. Denoising Diffusion Probabilistic Model[6], Score-based diffusion model[7] and the more recent Stable Diffusion model[8] , represent significant advancements

in this field Originally developed for generative tasks, diffusion models refine input data by progressively reconstructing it from a noisy, low-resolution state into a high-resolution output, effectively capturing intricate details often missed by traditional methods and frequently surpassing them in both quality and accuracy. The robust structure has allowed diffusion models to show much promising potential for super-resolution tasks as well, and is beginning to be applied to super-resolution tasks including meteorological data [9][10].

Despite their advantages, diffusion models still face challenges when applied to multimodal and fine-scale meteorological data. They often struggle to accurately capture the complex interactions between different types of weather data, such as solar irradiance, temperature, and wind speed. These limitations can lead to suboptimal performance in tasks that require precise high-resolution outputs.

To overcome these challenges, this paper proposes a diffusion-based framework specifically designed for the super-resolution of multimodal weather data. Our framework is based on the latest innovations in upsampling[11] and attention techniques[12] integrating the SwinT-DyS module to enhance the model's ability to process fine-scale data, ultimately improving the accuracy and reliability of PV potential estimation.

3. METHODOLOGY

3.1 SwinT-DyS

To improve the accuracy of meteorological data super-resolution within diffusion models, we initially employed the DySample module[11] as a dynamic upsampling solution within the Unet architecture. While DySample demonstrated significant advantages in computational efficiency and lightweight design, its performance at fine scales in meteorological data did not fully meet expectations. To address these limitations, we introduced the SwinT-DyS module, a fusion of the Swin Transformer[12] and DySample, aimed at enhancing the upsampling process for better results. The structure of the SwinT-DyS module is illustrated in Figure 1.

Figure 1 (a) shows the structure of two Swin Transformer consecutive blocks, each constructed based on shifted windows. These blocks differ from conventional multi-head self-attention (MSA) modules by employing a window-based mechanism. Each Swin Transformer block consists of LayerNorm (LN) layer, Multi-head Self Attention (MSA) module, residual connection, and 2-layer Multi-Layer Perceptron (MLP) with GELU non-linearity. The two consecutive Swin Transformer blocks alternate between Window-based

MSA (W-MSA) and Shifted Window-based MSA (SW-MSA). In the first block, W-MSA calculates attention within patches inside each window, focusing on localized regions. In the second block, SW-MSA shifts the windows in the top-left direction using a cyclic shifting mechanism, resulting in windows composed of non-adjacent sub-windows. The input and output of a continuous Swin Transformer block can be expressed as follows:

$$\hat{x}^l = W - MSA\left(LN(x^{l-1})\right) + x^{l-1} \qquad (1)$$

$$x^l = MLP\left(LN(\hat{x}^l)\right) + \hat{x}^l \qquad (2)$$

$$\hat{x}^{l+1} = SW - MSA\left(LN(x^l)\right) + x^l \qquad (3)$$

$$x^{l+1} = MLP\left(LN(\hat{x}^{l+1})\right) + \hat{x}^{l+1} \qquad (4)$$

where $\hat{x}^l$ and x^l represent the outputs of the (S)W-MSA module and the MLP module of the lth the block. The self-attention computed in W-MSA and SW-MSA can be written as follows:

$$Q = x^l W_Q, K = x^l W_K, V = x^l W_V, \qquad (5)$$

$$Attention(x^l) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \qquad (6)$$

where W_Q, W_K and $W_V \in \mathbb{R}^{D \times d}$ are the learnable parameters of three projection matrices. $Q, K, V \in \mathbb{R}^{L \times d}$ are the $query$, key, and $value$ matrices; d represents the dimension of $query$ or key. $B \in \mathbb{R}^{L \times d}$ denotes the relative position bias.
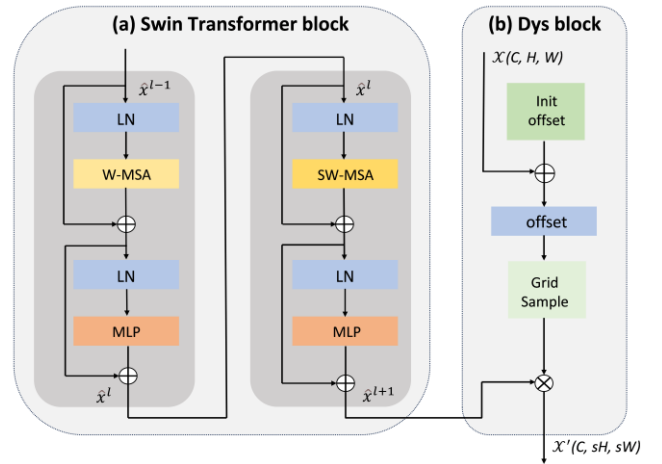

Fig 1. SwinT-DyS Module

The DySample module, as shown in Figure 1 (b), efficiently upscales feature maps using dynamic grid sampling. It begins by generating an offset map, which is combined with the original sampling grid to create a new sampling set. This set is then used to produce the upsampled feature map through bilinear interpolation.

To further enhance feature representation, we integrated the Swin Transformer into the DySample module. Specifically, the Swin Transformer is applied immediately after the dynamic sampling step within DySample. This integration allows the network to
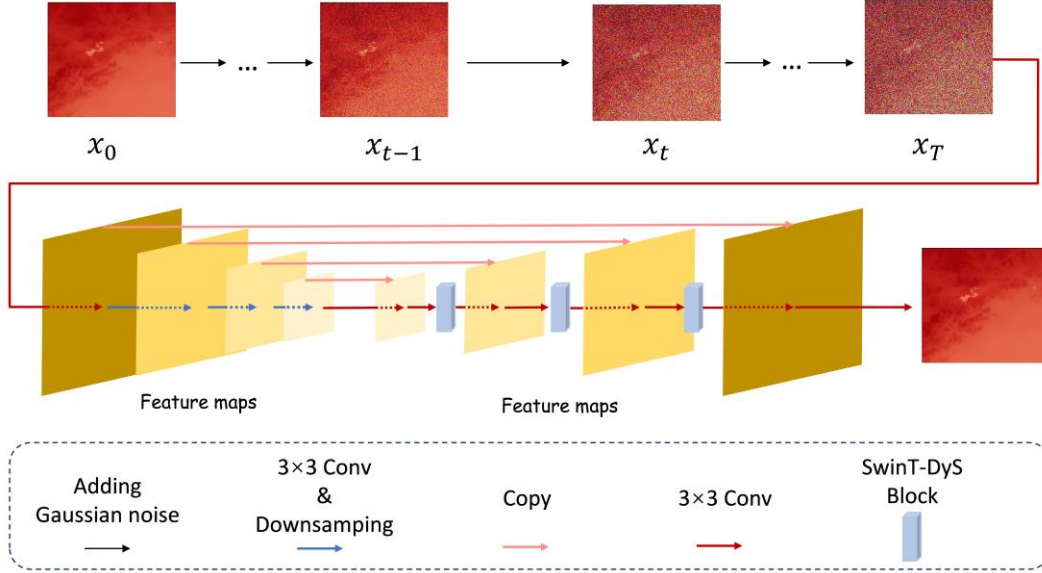
Fig 2. Diffusion Model with SwinT-DyS Block

process the upsampled feature maps with the Swin Transformer's window-based and shifted window-based multi-head self-attention (W-MSA and SW-MSA) mechanisms.

3.2 Diffusion model

We utilize a score-based diffusion model to enhance the resolution of multimodal meteorological data, specifically focusing on improving the accuracy of photovoltaic (PV) potential estimation. The core of our approach lies in a Score-Based Generative Model[7], which operates by iteratively refining noisy, low-resolution inputs to generate high-resolution outputs.

The forward diffusion process systematically corrupts the original low-resolution data x_0 by adding Gaussian noise over a time interval $t \in [0, T]$. This gradual noising process is described by the Stochastic Differential Equation (SDE):

$$dx = f(x, t)dt + g(t)dw \qquad (7)$$

where $f(x, t)$ denotes the drift coefficient, $g(t)$ is the diffusion coefficient responsible for the noise scale, and w represents a Wiener process. By the end of the forward process, the data is transformed into a standard Gaussian distribution, $x_T \sim \mathcal{N}(0, \mathrm{I})$.

To achieve super-resolution, the reverse diffusion process is employed to progressively denoise the corrupted data x_T, reconstructing a high-resolution version x_0. The reverse process is governed by the following SDE:

$$dx = [f(x, t) - g(t)^2 \nabla_x \log p_t(x)]dt + g(t)d\overline{w} \quad (8)$$

where $\overline{w}$ is a Wiener process with time flowing backwards from T to 0, $\nabla_x \log p_t(x)$ is the score function, a vector field that points towards regions of higher data density. It does not depend on the intractable normalization constant, making it easier to evaluate.

The score function $\nabla_x \log p_t(x)$ is parameterized by a neural network $s_\theta(x, t)$, which is implemented using a Unet architecture. The Unet is optimized to accurately estimate the score function across different noise levels by minimizing the following score-matching loss:

$$\theta^* = \arg \min_\theta \mathbb{E}_t\{\lambda(t)\mathbb{E}_{x(0)}\mathbb{E}_{x(t)|x(0)}[\| s_\theta(x, t) \qquad (9)$$
$$-\nabla_{x(t)} \log p_t(x(t)|x(0)) \|_2^2]\}$$

where $\lambda(t)$ is a weighting function that modulates the importance of different time steps during training. t is uniformly sampled from the interval $[0, T]$. This equation is derived by minimizing the evidence lower bound (ELBO) on the negative log-likelihood, $\mathbb{E}[-\log p(x_0)]$ with reweighting by $\lambda(t)$.

To further enhance the model's performance in super-resolving meteorological data, we integrate the SwinT-DyS module within the Unet's upsampling process. The SwinT-DyS module combines the robust attention mechanisms of the Swin Transformer with the dynamic sampling capabilities of DySample. This integration allows for more precise feature representation during the denoising process, thereby improving the overall accuracy of the super-resolved meteorological data and, consequently, the PV potential estimation.

4. EXPERIMENT

To validate the effectiveness of our diffusion model-based framework in enhancing the resolution of solar energy data, and to assess the significant benefits of the improved SwinT-DyS modules, we conducted a

comprehensive series of experiments. Model training was performed on a system equipped with a single NVIDIA RTX 4090D GPU (24GB) and a 16v Intel(R) Xeon(R) Platinum 8474C CPU, using a batch size of 8, a learning rate of 0.0001, and 40 epochs.

The experiments included a comparative analysis between the diffusion framework, a traditional interpolation method, and the well-established Unet model. Additionally, ablation studies were conducted to isolate and evaluate the specific contributions of the SwinT-DyS modules. To further demonstrate the impact of our super-resolution data, we conducted a PV Potential Estimation experiment to evaluate test accuracy, highlighting the role of enhanced data in improving photovoltaic potential predictions.

4.1 Datasets

For this study, the training data was sourced from the National Solar Radiation Database (NSRDB) for the Beijing region, using data from the years 2018 and 2019. For experimental validation, we utilized high-resolution data from the same region in January 2020, which also had a spatial resolution of 2 km, serving as our ground truth dataset. To create a low-resolution experimental dataset, we applied an interpolation method to downsample the 2020 high-resolution data, resulting in a dataset with a spatial resolution of 4km, 8km, 16 km. A detailed overview of the datasets used in this study is provided in the following table:

Characteristics	Values
Data feature	Temperature
	DHI
	DNI
	GHI
	Wind Speed
High Spatial Resolution	2km
Low Spatial Resolution	4km, 8km, 16km
Temporal Resolution	60 minutes
Model Name	PSM V3
Satellite	Himawari
Years	2018-2020

Table 1. Overview of the dataset

4.2 Evaluation metrics

In the experimental portion of this study, we evaluate the performance of our super-resolution framework using three widely recognized metrics: Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). These metrics provide a comprehensive assessment of the accuracy, quality, and perceptual similarity of the reconstructed high-resolution images compared to the ground truth.

MAE measures the average absolute difference between predicted and true values. The formula is:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i| \qquad (10)$$

where N is the total number of data points, x_i is the true value, and $\hat{x}_i$ is the predicted value.

PSNR quantifies the quality of the reconstructed image by comparing the maximum possible pixel value to the mean squared error (MSE), which measures the average of the squares of the differences between the true and predicted values. The formula is:

$$PSNR = 20 \cdot log_{10}\left(\frac{1}{\sqrt{MSE}}\right) \qquad (11)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \qquad (12)$$

where *N is the total number of data points, x_i is the true value, and $\hat{x}_i$ is the predicted value.*

SSIM assesses the perceptual similarity between two images, considering luminance, contrast, and structure. *The formula is:*

$$SSIM(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)} \qquad (13)$$

where μ_x and $\mu_{\hat{x}}$ are the mean intensities of the true and predicted images, σ_x^2 and $\sigma_{\hat{x}}^2$ are the variances of the true and predicted images, $\sigma_{x\hat{x}}$ is the covariance between the true and predicted images, C_1 and C_2 are stabilizing constants to avoid division by zero.

4.3 Results and discussion

4.3.1 Meteorological Data Super-Resolution

In this section, we present the experimental evaluation of three super-resolution methods: Bilinear interpolation, Unet, and Diffusion, applied to meteorological data from Beijing, January 2020. The experiment was conducted on five key meteorological variables: Temperature, Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), and Wind Speed with a spatial resolution of 16 km. The primary goal of this experiment is to enhance the spatial resolution of this data and compare the effectiveness of the three methods in terms of accuracy and quality across various meteorological variables.

As shown in Table 2, for the solar data variables (DHI, DNI, and GHI), the Diffusion model demonstrated a significant lead across all three measures (MAE, PSNR, and SSIM). This dominant performance is likely due to the model's strength in capturing and preserving the intricate spatial patterns and variability inherent in solar

Scale	Method	Temperature	DHI	DNI	GHI	Wind Speed
		MAE↓/PSNR↑/SSIM↑	MAE↓/PSNR↑/SSIM↑	MAE↓/PSNR↑/SSIM↑	MAE↓/PSNR↑/SSIM↑	MAE↓/PSNR↑/SSIM↑
×2	bilinear	**0.31/34.84/0.9680**	1.14/34.92/0.9919	7.09/30.35/0.9835	2.27/35.21/0.9942	**0.04/31.49/0.9903**
	Unet	0.43/33.54/0.9483	1.16/34.97/0.9918	7.09/30.37/0.9835	2.28/35.24/0.9940	0.42/21.30/0.6807
	Ours	0.41/34.28/0.9482	**1.05/35.41/0.9925**	**6.17/30.87/0.9862**	**2.03/35.68/0.9947**	0.37/24.63/0.6286
×4	bilinear	**0.45/33.16/0.9354**	1.57/33.40/0.9891	9.86/28.94/0.9742	3.13/33.82/0.9921	**0.05/31.40/0.9881**
	Unet	0.51/32.55/0.9232	1.59/33.45/0.9890	9.86/28.95/0.9743	3.14/33.85/0.9920	0.43/21.23/0.6687
	Ours	0.49/33.14/0.9310	**1.48/33.69/0.9894**	**8.88/29.39/0.9778**	**2.88/34.21/0.9927**	0.30/25.21/0.6268
×8	bilinear	**0.63**/31.28/0.8900	2.04/32.08/0.9851	12.96/27.59/0.9620	4.08/32.47/0.9895	**0.08/30.83/0.9785**
	Unet	0.67/30.78/0.8763	2.06/27.60/0.9620	12.99/27.60/0.9620	4.09/32.73/0.9893	0.46/21.29/0.9786
	Ours	0.64**/31.39/0.8998**	**1.97/32.21/0.9853**	**12.02/27.95/0.9660**	**3.85/32.78/0.9898**	0.32/24.85/0.6067

Table 2. Performance Comparison of Super-Resolution Methods Across Meteorological Variables at Different Scales

irradiance data. The Diffusion model's ability to maintain high structural similarity and low noise levels makes it particularly effective for these types of data. Among the solar variables, GHI exhibited the weakest performance, particularly in comparison to DHI and DNI. This could be attributed to the complex and diffuse nature of global horizontal irradiance, which combines both direct and scattered sunlight. The inherent variability and the diffuse component of GHI make it more challenging to model accurately, leading to slightly lower performance even with advanced techniques like the Diffusion model. A surprising finding from our experiments is that bilinear interpolation outperformed the deep learning method on the variables of Temperature and Wind Speed, especially at lower scales. This unexpected result suggests that there may be underlying factors related to the multimodal nature of the data or potential issues within the dataset itself. However, as the super-resolution scale increased, the gap between Bilinear interpolation and the more advanced methods narrowed. At the ×8 scale, the Diffusion model ultimately achieved superior results in PSNR and SSIM for both Temperature and Wind Speed. This shift suggests that the Diffusion model is better suited to handling the increased complexity and detail required at higher resolutions.

Future work could explore integrating more sophisticated loss functions tailored to the specific characteristics of different meteorological variables, potentially addressing the current limitations observed in GHI modeling. Additionally, further optimization of the Diffusion model to handle non-normally distributed data more effectively could make model even more robust across a wider range of applications The Diffusion model's ability to maintain high structural similarity and low noise levels makes it particularly effective for these types of data.

4.3.2   Ablation Study

The ablation study assesses the impact of integrating the SwinT-DyS module into the Diffusion model on super-resolution performance, using DHI data from January 2020 for comparison. As shown in Table 3, the Diffusion model with the integrated SwinT-DyS module achieves a lower MAE, indicating a reduction in absolute errors, while the higher PSNR and SSIM values demonstrate improvements in image quality and structural preservation. The SwinT-DyS module clearly plays a critical role in enhancing the model's overall performance, making it more effective for super-resolution tasks involving meteorological data.

Diffusion	SwinT-DyS	MAE↓	PSNR↑	SSIM↑
✓		1.17	34.93	0.9919
✓	✓	**1.05**	**35.41**	**0.9925**

Table 3. Ablation Study Results

4.3.3   Enhanced PV Potential Estimation Using Diffusion-Based Super-Resolution Data

In this study, we employ the pvlib library[13] to estimate the photovoltaic (PV) potential using meteorological data at different spatial resolutions for Beijing in January 2020. The primary objective is to compare the PV potential estimates derived from low-resolution data against those generated using super-resolution data produced by our diffusion model.

The analog PV system is configured with a fixed tilt angle of 40° and oriented due south, which is typical for installations in the Beijing region. The system parameters are modeled using the Sandia PV module and CEC inverter models provided within pvlib. The calculation involves determining the total plane-of-array (POA) irradiance, considering both direct and diffuse components, as well as the module's cell temperature, based on the relevant data used to produce the super-resolution output. The DC power output is computed based on the effective irradiance and module temperature, then converted to AC power using the inverter model. The resulting ac power output is
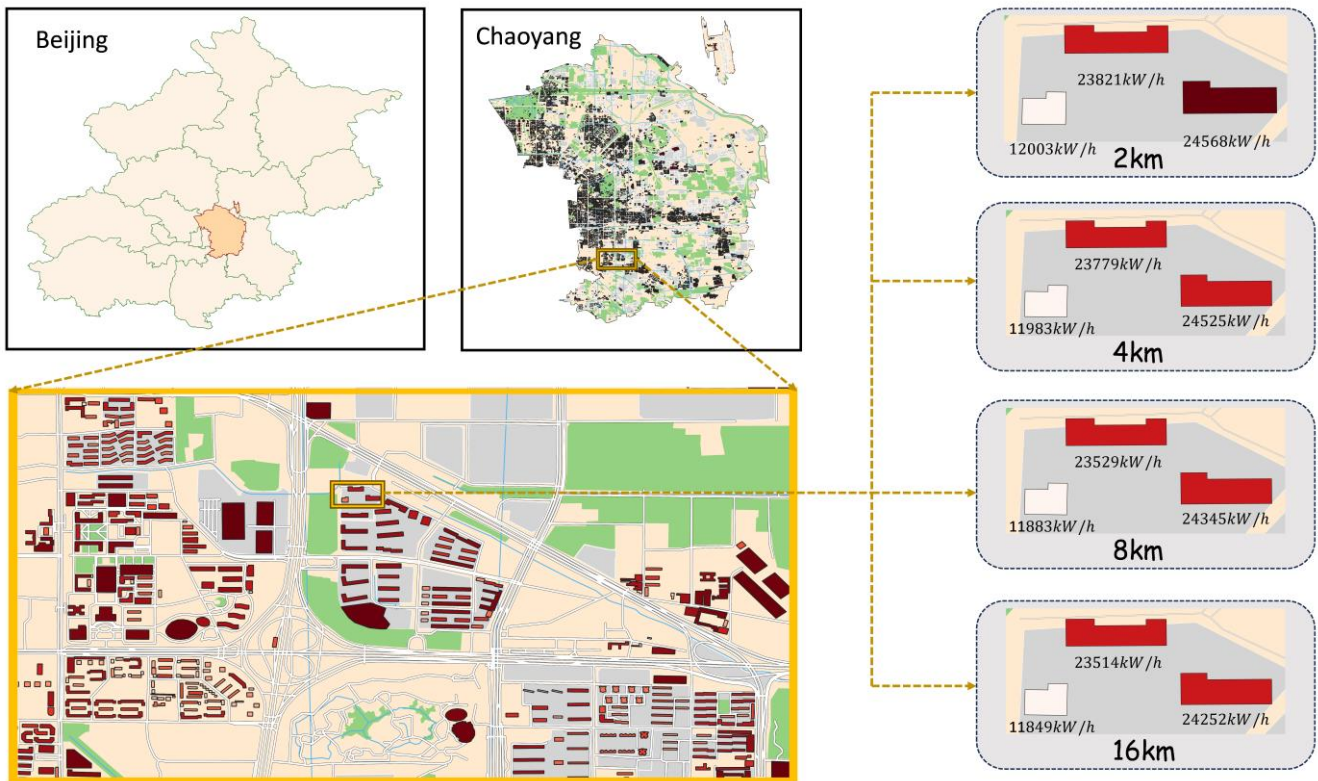
Fig 3. Chaoyang Rooftop PV Potential Comparison High vs Low Resolution. The left side of the figure shows an example of PV output visualization for part of Chaoyang, Beijing. The right side highlights the differences in PV output when comparing the high-resolution data with a coarser resolution data.

aggregated over time to estimate the total PV generation potential of each grid point for a whole month.

We then integrated the data on PV potential per unit area with the data on roof area in Beijing to calculate the PV potential of each roof, as seen in Figure 4 for the visualization of rooftop PV potential in Beijing. After visualization, it becomes evident that the estimates from the high-resolution data offer several advantages. The finer spatial resolution provides a more detailed and accurate depiction of the PV potential across the region, capturing subtle variations that are missed in the low-resolution data. This enhanced clarity in the high-resolution results underscores the value of super-resolution techniques, which significantly improve the precision of renewable energy assessments. These advantages highlight the potential of our diffusion-based approach to refine energy modeling and planning, making it an essential tool for future applications.

5. CONCLUSIONS

This study has demonstrated the effectiveness of a diffusion model-based framework in enhancing the resolution of multimodal meteorological data, leading to more accurate photovoltaic potential estimations. By integrating advanced techniques such as the SwinT-DyS module within the Unet architecture, our approach significantly improves the quality of super-resolution outputs. The results underscore the importance of high-resolution data in renewable energy applications and highlight the potential of our method to support more precise and reliable energy assessments, thereby contributing to the optimization of solar power deployment strategies.

REFERENCE

[1] Gassar, A. A. A., & Cha, S. H. (2021). Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. Applied Energy, 291, 116817.

[2] Chen, Q., Li, X., Zhang, Z., Zhou, C., Guo, Z., Liu, Z., & Zhang, H. (2023). Remote sensing of photovoltaic scenarios: Techniques, applications and future directions. Applied Energy, 333, 120579.

[3] Guo, Z., Lu, J., Chen, Q., Liu, Z., Song, C., Tan, H., ... & Yan, J. (2024). TransPV: Refining photovoltaic panel detection accuracy through a vision transformer-based deep learning model. *Applied Energy*, *355*, 122282.

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144.

[5] Cai, Z., Xiong, Z., Xu, H., Wang, P., Li, W., & Pan, Y. (2021). Generative adversarial networks: A survey toward private and secure applications. *ACM Computing Surveys (CSUR)*, *54*(6), 1-38.

[6] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

[7] Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456.

[8] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, *1*(2), 3.

[9] Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., ... & Chen, Y. (2022). Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing, 479, 47-59.

[10] Hatanaka, Y., Glaser, Y., Galgon, G., Torri, G., & Sadowski, P. (2023). Diffusion models for high-resolution solar forecasts. *arXiv preprint arXiv:2302.00170*.

[11] Liu, W., Lu, H., Fu, H., & Cao, Z. (2023). Learning to upsample by learning to sample. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6027-6037).

[12] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).

[13] Anderson, K., Hansen, C., Holmgren, W., Jensen, A., Mikofski, M., and Driesse, A. (2023). pvlib python: 2023 project update. Journal of Open Source Software, 8(92), 5994.