Quantized SDiT with Non-structured Positional Embedding for Signal Synthesis

Yunhan Xing[1], Yidong Li[1], Zikai Zhang[2*], Khaled A. Harras[2]
[1]Beijing Jiaotong University, China. [2]Carnegie Mellon University, Qatar.

***Abstract*—Synthesizing high-fidelity wireless signals is essential for addressing data scarcity, yet conventional generative models often fail to account for the non-structured spatial distribution of signal samples and suffer from excessive computational demands. This paper presents SDiT, a quantized Diffusion Transformer for signal synthesis. To address the lack of local spatial correlation in signal sequences, we modify the standard DiT embedding layer by concatenating learnable position embedding into samples and incorporating Rotary Position Encoding to embed continuous spatial coordinates into condition vector. By aligning feature representations with the underlying physical laws of signal propagation, this approach effectively captures the non-structured spatial dependencies of signal samples, successfully transcending the limitations inherent in grid-based image priors. Furthermore, to facilitate efficient deployment, we develop a block-wise quantization algorithm featuring an inter-channel covariance compensation mechanism. By minimizing the Mean Squared Error of layer outputs, this approach effectively suppresses quantization noise accumulation across denoising timesteps. Evaluations on WiFi, FMCW, and MIMO datasets demonstrate that our method significantly reduces model bit-width while maintaining superior generation metrics, including FID and SNR, proving its effectiveness for resource-constrained signal simulation tasks.**

***Index Terms*—Signal Synthesis, Diffusion Transformer, Position Embedding Map, Post-training Quantization.**

I. Introduction

The rapid advancement of wireless communication technologies has led to an increasing demand for high-fidelity signal samples to support critical tasks such as environment sensing [1] and network optimization [2]. A prominent example is accurate and ubiquitous floor identification [3], where synthesizing diverse signal patterns under the single cell tower (unicellular) constraint is essential for training robust localization models without relying on multi-tower infrastructure. However, collecting large-scale, high-quality signal data in real-world scenarios remains challenging [4]. While deep generative models has demonstrated remarkable success in data synthesis [5], its application to wireless signals is still hindered by the mismatch between signal physics and standard architectures, as well as the prohibitive computational and memory demands that exceed the limits of typical edge-deployed systems.

A fundamental obstacle in synthesizing high-information-value signal data lies in the structural mismatch between wireless signal physics and conventional generative architectures. As illustrated in Figure 1, unlike natural images that possess inherent local correlations within a rigid pixel grid, wireless signals are typically characterized by non-structured spatial distributions, where signal strength measurements are tied to physical coordinates. Image-based diffusion models, primarily optimized for grid-based data priors, impose artificial spatial constraints when forced to process signal samples as image-like patches [5]. Such architectures fail to capture the underlying dependencies governed by physical propagation laws, often resulting in synthesized samples that are visually plausible but physically inconsistent.

Beyond architectural alignment, the practical viability of diffusion model for wireless signal synthesis is dictated by its efficiency on resource-constrained edge hardware. While diffusion models excel at capturing complex distributions, their recursive denoising nature imposes a heavy computational burden that necessitates aggressive model compression [6]. However, standard post-training quantization (PTQ) often leads to catastrophic fidelity loss in diffusion frameworks [6]. This degradation stems from the quantization on weight outliers magnifies the error and the dynamic shift of activation distributions across different timesteps. Drawing inspiration from recent advances in block-wise optimization [7], we posit that block-wise quantization with compensating residual weights can suppress error accumulation within each linear layers.

To address the aforementioned challenges, we propose Signal Diffusion Transformer (SDiT), a novel generative framework engineered to synthesize physically-consistent wireless signals while maintaining extreme efficiency for edge deployment. Specifically, SDiT reconfigures the standard Transformer embedding layer by concatenating learnable positional embeddings with raw samples and utilizing Rotary Position Encoding to project continuous spatial coordinates into the condition vector. This hybrid embedding strategy allows the model to align its latent representations with the non-structured nature of signal propagation laws rather than rigid image-centric grids. Furthermore, we introduce a specialized block-wise quantization algorithm featuring an inter-channel covariance compensation mechanism. By formulating the quantization process as a layer-wise optimization problem, our method

Zikai Zhang is the corresponding author(zikaiz2@cmu.edu). This work was supported in part by ARG02-0429-240389 from the Qatar National Research Fund/Qatar Research, Development and Innovation Council; the National Science Foundation of China under Grant U2268203, 62306027 and 62372436; the Shanxi Province's Major Science and Technology Special Program "Unveiling the List and Leading the Way" Project (Grant No. 202401050202010); and the Open Fund of State Key Laboratory of Digital Intelligent Technology for Unmanned Coal Mining (Grant No.SZQZ2025-2-004).

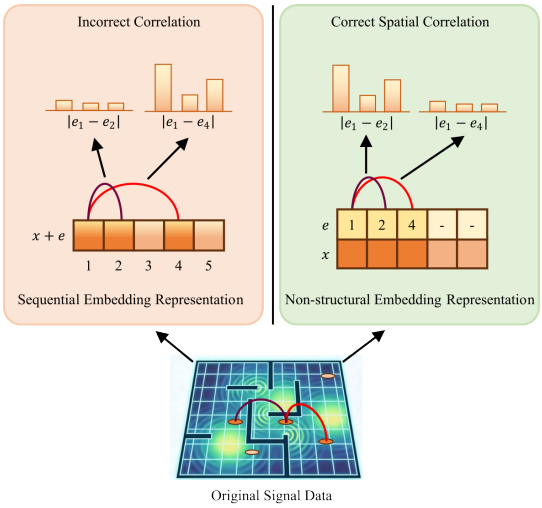

Fig. 1. Diagram of signal data into sequential embedding in DiT and SDiT. To avoid the incorrect local correlations imposed by grid-based DiT priors on non-uniform wireless samples, SDiT leverages non-structural embedding representation to capture correct intrinsic spatial dependencies.

minimizes the Mean Squared Error (MSE) of feature reconstructions, effectively suppressing the cumulative noise inherent in the recursive denoising process. Experimental results across multiple signal modalities, including WiFi, FMCW, and MIMO, demonstrate that SDiT achieves state-of-the-art synthesis fidelity while enabling significant bit-width reduction for resource-constrained tasks.

The main contributions are as follows,

- We propose SDiT, a specialized Diffusion Transformer that transcends grid-based image constraints by incorporating a dual-embedding mechanism. This allows for the precise capture of spatial dependencies in non-structured signal samples by aligning with physical propagation laws.
- We develop a block-wise post-training quantization algorithm specifically optimized for the dynamic activation shifts in diffusion models. By introducing inter-channel covariance compensation, we effectively minimize MSE during the denoising process, preserving signal integrity (SNR) at low bit-width.
- We validate our framework on diverse wireless datasets, demonstrating its superior capability in generating high-information-value samples. Our results confirm that SDiT provides a robust and efficient solution for data augmentation in edge-deployed sensing tasks.

II. Background and Related Works

A. Signal Synthesis

In signal processing fields, the performance of data-driven deep learning models heavily relies on large-scale, high-quality training data. However, in practical scenarios, acquiring sufficient and diverse real signal data is often challenging due to high collection costs, privacy constraints, or the difficulty of reproducing specific rare scenarios. This data bottleneck directly limits the performance and generalization capability of downstream tasks such as channel estimation [8], target localization [9], and anomaly detection [10].

Effective signal synthesis is essential to address data scarcity, yet conventional generative models, including GANs [11] and AEs [12], often fail to generalize across complex signal distributions. While diffusion models offer superior generation quality [13], their recursive denoising nature imposes prohibitive computational costs, necessitating efficient architectural and quantization strategies for edge deployment.

B. Diffusion Transformer

Recently, DiT [13] have achieved revolutionary success in the field of generative artificial intelligence, rapidly becoming the de facto standard for multimedia content generation tasks such as image synthesis [14], video generation [15], and audio production [16]. Compared to U-Net based model, DiT demonstrates better scalability.

Research on applying DiT to the field of wireless signal synthesis is still in its nascent stage. Wireless signal data are typically composed of multi-dimensional heterogeneous attributes and lack inherent spatial locality. This fundamentally limits the direct transfer and application of traditional visual diffusion models. Among the few explorations, RF-Diffusion [17] is a pioneering work that first introduced diffusion models to the field of radio frequency signal generation, employing complex-valued transformations to handle signal phase and amplitude information. However, this work and subsequent studies still primarily rely on model architectures designed for images, which are not optimized for signal data.

C. Diffusion Model Quantization

PTQ [18] is widely adopted to enhance inference efficiency and reduce memory footprint. Given the iterative nature of diffusion models, PTQ effectively alleviates their heavy computational overhead. Current research, such as PTQ4DiT [19], focuses on mitigating extreme activation distributions through channel-wise balancing, while Q-DiT [20] employs sample-wise dynamic quantization to handle activation variations across timesteps. However, these methods primarily address marginal distribution statistics and often overlook the inter-channel error correlations during the recursive denoising process, which can lead to catastrophic noise accumulation at low bit-widths. In contrast, our proposed SDiT introduces an inter-channel covariance compensation mechanism that minimizes layer-wise reconstruction error, ensuring high signal fidelity even under aggressive quantization without the heavy overhead of dynamic schemes.

III. Preliminary

A. Diffusion Model

The diffusion model utilize inverse of propagating process to generate samples. A forward propagation process can be described as a Markov chain which formulate as:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t} x_{t-1}, \sqrt{1-\alpha_t} I), \quad (1)$$

where α_t represent noise schedule. Sample $x_0 \in \mathbb{R}^n$, which satisfies signal embedding distribution, gradually propagating into Gaussian noise, where $t \in \{1, \ldots, T\}$ denotes propagation timestep.

In backward process, diffusion model employs probabilistic network to inference original samples:

$$p_\theta\left(x_{t-1} \mid x_t\right) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \tag{2}$$

In Denoising Diffusion Implicit Model (DDIM) [21], the mean value of the backward distribution can be computed by a noise prediction network $\epsilon_\theta(x_t, t)$ through Eq. (3) and the variance is reparameterized into a fixed schedule $\sigma_t \boldsymbol{I}$.

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) \tag{3}$$

To train denoising network, we use KL divergence as loss function, which is proportional to MSE loss between real noise and predicate result from denoising model:

$$D_{KL}\left(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t)\right) \propto \left\|\epsilon - \epsilon_\theta(x_t, t)\right\|^2. \tag{4}$$

B. Post-training quantization

Post-training quantization (PTQ) fine-tune model parameters into lower bit-width through quantize function after model training. We use n-bit uniform quantization function to quantize weights and activation:

$$Q(x; s, z) = s \cdot \left(\left\lceil \mathrm{clip}\left(\frac{x}{s} - z, 0, 2^n - 1\right)\right\rfloor + z\right) \tag{5}$$

where s and z are quantization parameters. $\lceil \cdot \rfloor$ denotes rounding function and $\mathrm{clip}(\cdot)$ represents clipping function.

C. Problem Statement

Given a labeled radio signal dataset $\mathcal{D}$. Each element $(\boldsymbol{x}, \boldsymbol{c}) \in \mathcal{D}$ contains radio signal strength $\boldsymbol{x} \in \mathbb{R}^n$ (RSS) received from each transistor, and environment label $\boldsymbol{c} \in \mathbb{R}^{n_c}$ for downstream tasks. Our problem is designing and training a diffusion model $f(\boldsymbol{x}, \boldsymbol{c})$ to generate synthetic data $\boldsymbol{y} \in \mathbb{R}^n$ of which distribution is close to the realistic distribution of data. Specifically, we need design a input encoding $f_x \in \mathbb{R}^n \to \mathbb{R}^{h_x}$ and condition embedding module $f_c \in \mathbb{R}^{n_c} \to \mathbb{R}^{h_c}$ which has better performance, and quantize model into lower bit width with retaining the generation quality.

IV. Method

A. Design of Diffusion Transformer Architecture

Unlike image features, radio signals comprise attributes such as signal strength, location, and other metadata. This presents two key challenges for network architecture design: first, the condition input requires a composite embedding that integrates both continuous and discrete elements; second, signal strength is typically represented as an unstructured sequential data that does not possess the spatial locality characteristic of image data.

To address these challenges, we proposed a SDiT denoising framework, processing signals through an input projection layer, a condition embedding layer, multiple transformer blocks, and an output projection layer to reconcile discrete signal attributes with continuous spatial conditions.

For continuous features such as coordinates, environmental obstructions cause signal strength to be discontinuous with respect to location. Directly using raw coordinates for generation thus leads to poor quality. To address this challenges, inspired by Implicit Neural Representations [22], SDiT models the radio environment as a continuous Wave Field. By bypassing discrete grid priors for continuous coordinate embeddings, our framework implicitly learns to satisfy the Electromagnetic Wave Equation [23]. This enables the capture of complex interference patterns and spatial correlations. Consequently, we employ a contiguous form of rotary position encoding:

$$h_{uv} = \exp\left(i\omega_v p_u\right) \tag{6}$$

where p_u is the u-th component of coordinate $\boldsymbol{p}$ and ω_v is the rotary scale for the v-th channel.

This technique maps coordinates into a high-dimensional manifold to preserve relative spatial dependencies. The resulting features are fused with learnable categorical embeddings (e.g., environment types) and timestep information, then injected into the Transformer blocks via an adaptive layer normalization (adaLN) interface [24].

Departing from the standard DiT architecture, we adapt the input encoding layer within the DDIM framework to incorporate explicit local structures, as the backbone for non-structured signal synthesis. We design an independent feature embedding vector, as shown in Fig. 2, for each signal source within the data sequence. This conditional encoding is then concatenated and fused with the original noisy signal sample along the feature dimension, collectively forming the input sequence for the denoising transformer. This design enables the model to directly process the unstructured intensity sequence of the signals while associating it with corresponding source-level conditional features, thereby learning the conditional distribution for signal generation more effectively.

B. Model Quantization

To reduce the computational and storage costs of the trained denoising diffusion model at inference, we apply PTQ to both weights and activations while aiming to control the quantization error, as shown in Algorithm 1. The specific procedure is as follows:

First, a calibration dataset is generated with the full-precision model to capture the activation distributions. Subsequently, by independently analyzing input channels and dynamically adjusting quantization parameters based on activation distributions, quantization is applied to the weights to obtain an initial quantized model. Finally, using the calibration data, the dynamic ranges of activations are analyzed to determine and calibrate layer-wise quantization parameters, yielding a fully quantized low-bit model.

To mitigate errors from independent per-channel quantization, we introduce an inter-channel covariance compensation mechanism. For a linear layer, we decompose the output Y to isolate the i-th quantized channel, allowing us to adjust

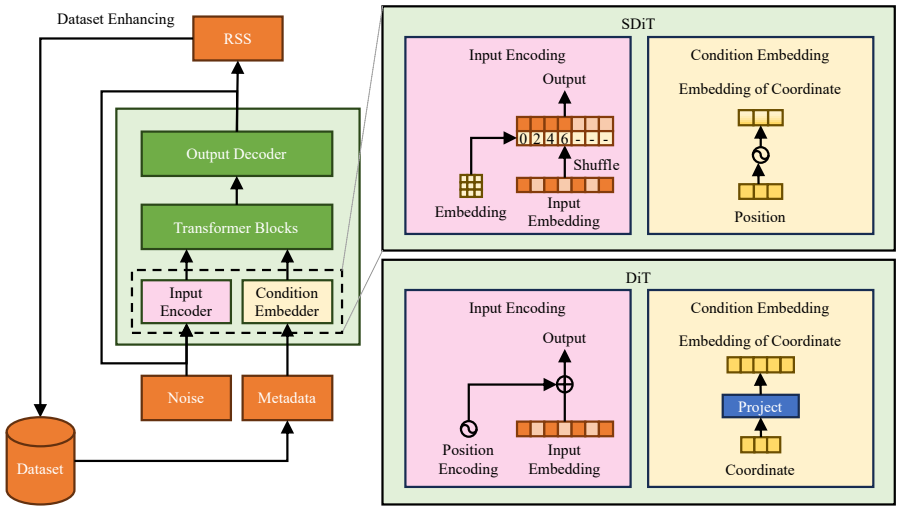

Fig. 2. SDiT architecture design and comparison of its input layer with DiT model.

Algorithm 1 Weight Quantization for SDiT

Input: The FP network ϵ_θ.
Input: The calibration dataset X.
Output: The quantized network $\epsilon_{\hat{\theta}}$.

```
Initialize weight quantized parameters.
for k-th linear layer to be quantized in network do
    Compute μ = h̄, Σ = h^⊤ h, where h denote layer input evaluated from dataset X.
    for i-th input channel for k-th linear layer do
        Quantize weight w_i.
        Update residual full-precision weight W_i ← W_i − ΔŴ_i(μ_i, Σ_i).
    end for
end for
Quantization activation of network ε_θ.
```

its parameters by accounting for its covariance with all other activation channels. The equation is:

$$Y = X_i W_i + x_i Q(w_i), \tag{7}$$

where x_i and w_i are the input and full-precision weight vector for the i-th channel, respectively, and $Q(w_i)$ is its quantized counterpart. X_i and W_i denote the concatenated inputs and weights for all other channels.

The quantization error for this layer is measured by the MSE of the output perturbation. This perturbation ΔY arises from replacing the full-precision weight w_i with its quantized version $Q(w_i)$, while compensating for the introduced error by jointly updating the weights of other channels W_i:

$$\Delta Y = X_i \Delta W_i + x_i \Delta w_i, \tag{8}$$

where $\Delta w_i = Q(w_i) - w_i$ is the intrinsic quantization error of the i-th channel's weights, and ΔW_i is the compensatory update applied to other channels' weights. Our objective is to find the optimal compensation ΔW_i that minimizes the output MSE:

$$\Delta \hat{W}_i = \underset{\Delta W_i}{\operatorname{argmin}} \, (\Delta Y)^2. \tag{9}$$

To simplify the loss function, we model the linear dependence between the input of the i-th channel x_i and the inputs of all other channels X_i as:

$$x_i = X_i u_i^\top + \epsilon(v_i), \tag{10}$$

where u_i is a coefficient vector, and $\epsilon(v_i)$ is a residual term with a constant mean $\mathbb{E}[\epsilon(v_i)] = v_i$ and is uncorrelated with X_i.

Substituting this relation into Eq. (8) and minimizing the quadratic form $\mathbb{E}[(\Delta Y)^2]$, we obtain ΔW_i that yields the optimal compensatory weight update:

$$\Delta \hat{W}_i = -\frac{1}{2}(\Sigma_i u_i \Delta w_i^\top + v_i \Delta w_i \mu_i^\top)\Sigma_i^{-1}, \tag{11}$$

where $\Sigma_i = \mathbb{E}[X_i^\top X_i]$ and $\mu_i = \mathbb{E}[X_i]$ are the covariance matrix and mean vector of X_i.

V. Implementation Details

Our model is implemented using PyTorch and the Diffusers library, and all experiments are conducted on NVIDIA L20 GPUs. For the diffusion process, we employ a linear noise schedule with $T = 1000$ timesteps, where the noise variance parameters are set to $\beta_T = 2 \times 10^{-2}$ and $\beta_0 = 1 \times 10^{-4}$, following the definition $\beta_t = \sqrt{1 - \alpha_t}$.

To facilitate stable training and quantization, both the signal strength and coordinate values in the dataset are normalized to the range $[-1, 1]$. Furthermore, to ensure a consistent input sequence length for the transformer, all RSS sequences are padded to the length of the longest sequence in the dataset using a minimal signal strength value as the padding token.

The model is optimized using the AdamW optimizer with an initial learning rate of 1.0×10^{-3}. A step decay learning rate scheduler with a decay rate of 0.8 is applied during

training. In result, quantization settings WxAy denotes weight in x bits and activation in y bits, and Linear represent linear quantization as baseline of quantization method.

VI. Experiments

A. Evaluation of Signal Generation Quality and Efficiency

a) Objective and Metrics: This experiment aims to evaluate the impact of our proposed architectural improvements and quantization methods. Specifically, we assess whether the model maintains or even improves generation quality while achieving significant reductions in computational cost. The evaluation employs the following metrics:

- **Structural Similarity Index Measure (SSIM)** [25]: Quantifies the structural similarity between two samples by comparing their luminance, contrast, and structure.
- **Fréchet Inception Distance (FID)** [26]: Measures the similarity between the distributions of real and generated samples using features extracted by a pre-trained network.
- **Signal-to-Noise Ratio (SNR)**: Assesses signal quality by calculating the ratio of signal power to noise power.
- **Bit Operations (BOPs)**: Estimates the computational cost of the quantized model by measuring the total bits processed during inference.

b) Datasets and Experimental Setup: We conduct evaluations on three datasets: the WiFi, FMCW, and MIMO datasets from [17]. For a fair comparison, the base SDiT model is trained on data synthesized by the HDT model [17], which also serves as our primary baseline.

Additionally, we evaluate performance on the *Unicellular* dataset [27], which contains 19,600 samples of Received Signal Strength (RSS) vectors collected across 9 floors. We adopt a 70%/30% train-test split. To complement the standard metrics, we introduce a downstream task-based fidelity assessment: a classifier is trained on the *real* training subset to predict floor labels, and its accuracy drop when applied to generated samples (versus the held-out real test set) quantifies generation quality.

c) Results and Analysis: The results for standard signal generation are summarized in Table I. On the WiFi and FMCW datasets, our SDiT model achieves an SSIM of 0.81 and an FID of 5.68. While the FID shows a slight increase compared to the HDT baseline, the SSIM remains high. Crucially, the quantized version delivers nearly identical SSIM performance, demonstrating the effectiveness of our quantization method.

For the MIMO channel estimation task (Table II), our model attains an SNR of 29.13 dB, a marginal decrease of only 0.82 dB from the HDT baseline, confirming its robustness under quantization.

Results on the Unicellular dataset (Table III) further validate our design. The full-precision SDiT achieves a classification accuracy of 77.1%, outperforming the baseline DiT by 4.6%, which highlights the benefit of our architectural modifications. The quantized model exhibits only a minor accuracy degradation, underscoring its practical utility.

TABLE I
Result of WiFi and FMCW Signal Generation

Dataset	Model	FID↓	SSIM↑
WiFi	HDT	5.64	0.81
	SDiT	5.68	0.81
	SDiT(Linear, W8A8)	5.24	0.81
	SDiT(ours, W8A8)	5.08	0.81
	SDiT(Linear, W4A8)	7.13	0.80
	SDiT(ours, W4A8)	6.73	0.80
FMCW	HDT	5.70	0.76
	SDiT	5.68	0.76
	SDiT(Linear, W8A8)	5.31	0.75
	SDiT(ours, W8A8)	5.37	0.77
	SDiT(Linear, W4A8)	6.12	0.75
	SDiT(ours, W4A8)	5.93	0.75

TABLE II
Result of MIMO Signal Generation

Dataset	Model	SNR↑
MIMO	HDT	29.95
	SDiT	29.13
	SDiT(Linear, W8A8)	28.36
	SDiT(ours, W8A8)	28.48
	SDiT(Linear, W4A8)	22.29
	SDiT(ours, W4A8)	24.62

B. Evaluation of Data Augmentation Utility

We conduct data augmenting experiment on the Unicellular dataset with the floor classification task.

The goal is to verify whether augmenting with synthetic samples can maintain model performance while reducing dependency on real data. Fully-connected networks are trained on datasets augmented with varying proportions of synthetic data, and their classification accuracy is reported.

As shown in Figure 3, our SDiT-based augmentation strategy consistently outperforms baseline methods across all tested data substitution rates. Specifically, the SDiT increases 1.2% accuracy when the model is trained with 90% generated data, and the quantized model have only 0.6% accuracy drop at most among all substitution rates. This indicates that our generated samples not only possess high fidelity but also effectively enhance the diversity and representativeness of the training data, leading to more robust downstream models.

VII. Conclusion

This paper introduced SDiT, a Signal Diffusion Transformer framework for high-fidelity wireless signal synthesis. By modifying the embedding layer through Rotary Position Encoding, SDiT successfully combine the non-structured spatial distribution of signal samples with the generative power of Transformers. This design bypasses the grid-centric constraints typical of image-based priors, enabling the latent feature space to directly align with the physical laws of signal propagation. To address edge deployment constraints, we further introduce a block-wise quantization algorithm incorporating inter-channel covariance compensation. By minimizing layer-wise

TABLE III
RESULT OF DATA GENERATION FOR UNICELLULAR DATASET

Model	BOPs	FID↓	Accuracy↑
Real	-	-	85.1%
VAE	6.63B	18.67	23.8%
GAN-VAE	6.63B	0.18	48.5%
DiT	494.7B	0.08	72.5%
SDiT	494.7B	0.06	77.1%
SDiT(W8A8)	127.2B	0.08	75.9%
SDiT(W4A8)	127.2B	0.08	77.7%

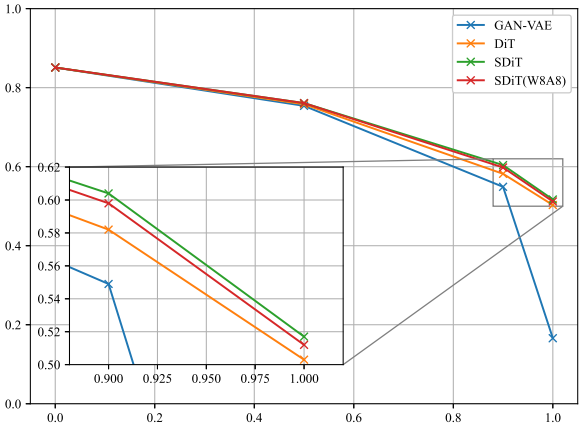

Fig. 3. Result of Unicellular Dataset Augmentation

reconstruction error (MSE), the proposed quantization strategy curtails noise accumulation across the iterative denoising trajectory, preserving signal integrity even at aggressive bit-widths. Extensive benchmarks on WiFi, FMCW and MIMO modalities demonstrate that Quantized SDiT yields a superior balance of synthesis fidelity (SNR/FID) and computational efficiency, serving as a practical solution for addressing data scarcity in real-world wireless sensing tasks.

REFERENCES

[1] H. Nasiri, S. Dogan-Tusha, M. I. Rochman, and M. Ghosh, "Data driven environmental awareness using wireless signals," *arXiv preprint arXiv:2410.13159*, 2024.

[2] M. Subramaniyan, T. K. Venkatasamy, N. Thotakuruchi Mathiyalagan, and M. Jakir Hossen, "Adaptive resource allocation and routing for integrated sensing and communications for wireless technologies," *EURASIP Journal on Wireless Communications and Networking*, vol. 2025, no. 1, p. 33, 2025.

[3] S. Mostafa, M. Youssef, and K. A. Harras, "Ubiquitous and low-overhead floor identification with limited cellular information," *ACM Transactions on Spatial Algorithms and Systems*, 2024.

[4] J. Peng, Z. Chen, Z. Lin, H. Yuan, Z. Fang, L. Bao, Z. Song, Y. Li, J. Ren, and Y. Gao, "Sums: Sniffing unknown multiband signals under low sampling rates," *IEEE Transactions on Mobile Computing*, 2024.

[5] U. Sengupta, C. Jao, A. Bernacchia, S. Vakili, and D. Shiu, "Generative diffusion models for radio wireless channel modelling and sampling," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 4779–4784.

[6] X. Li, Y. Liu, L. Lian, H. Yang, Z. Dong, D. Kang, S. Zhang, and K. Keutzer, "Q-diffusion: Quantizing diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 535–17 545.

[7] E. Frantar, S. Ashkboos, T. Hoefler, and D. Alistarh, "Gptq: Accurate post-training quantization for generative pre-trained transformers," *arXiv preprint arXiv:2210.17323*, 2022.

[8] F. Pasic, L. Eller, S. Schwarz, M. Rupp, and C. F. Mecklenbräuker, "Deep learning-based mmwave mimo channel estimation using sub-6 ghz channel information: Cnn and unet approaches," *arXiv preprint arXiv:2506.11714*, 2025.

[9] Y. Zook, A. Shokry, and M. Youssef, "An efficient quantum binary-neuron algorithm for accurate multi-story floor localization," in *2024 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2024, pp. 1–6.

[10] X. Liu, W. Tan, Z. Li, and J. Zeng, "Unsupervised dl-based wireless signal anomaly detection," *Digital Signal Processing*, p. 105578, 2025.

[11] M. Thirunavukkarasu, S. Kuzhaloli, S. Sugumaran, T. Lakshmibai, T. D. Kumar, and P. Karthick, "Generating synthetic csi data using gans: a deep learning-based approach," in *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*. IEEE, 2025, pp. 276–283.

[12] L. Li, Z. Zhang, and L. Yang, "Influence of autoencoder-based data augmentation on deep learning-based wireless communication," *IEEE Wireless Communications Letters*, vol. 10, no. 9, pp. 2090–2093, 2021.

[13] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4195–4205.

[14] Y. Yu, W. Xiong, W. Nie, Y. Sheng, S. Liu, and J. Luo, "Pixeldit: Pixel diffusion transformers for image generation," *arXiv preprint arXiv:2511.20645*, 2025.

[15] G. Chen, D. Lin, J. Yang, C. Lin, J. Zhu, M. Fan, H. Zhang, S. Chen, Z. Chen, C. Ma *et al.*, "Skyreels-v2: Infinite-length film generative model," *arXiv preprint arXiv:2504.13074*, 2025.

[16] Y. Jia, Y. Chen, J. Zhao, S. Zhao, W. Zeng, Y. Chen, and Y. Qin, "Audioeditor: A training-free diffusion-based audio editing framework," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[17] G. Chi, Z. Yang, C. Wu, J. Xu, Y. Gao, Y. Liu, and T. X. Han, "Rf-diffusion: Radio signal generation via time-frequency diffusion," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, 2024, pp. 77–92.

[18] M. Nagel, M. Fournarakis, R. A. Amjad, Y. Bondarenko, M. Van Baalen, and T. Blankevoort, "A white paper on neural network quantization," *arXiv preprint arXiv:2106.08295*, 2021.

[19] J. Wu, H. Wang, Y. Shang, M. Shah, and Y. Yan, "Ptq4dit: Post-training quantization for diffusion transformers," *Advances in neural information processing systems*, vol. 37, pp. 62 732–62 755, 2024.

[20] L. Chen, Y. Meng, C. Tang, X. Ma, J. Jiang, X. Wang, Z. Wang, and W. Zhu, "Q-dit: Accurate post-training quantization for diffusion transformers," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 28 306–28 315.

[21] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[22] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[23] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[24] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[26] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.

[27] S. Mostafa, M. Youssef, and K. A. Harras, "Accurate and ubiquitous floor identification at the edge using a single cell tower," in *2024 IEEE/ACM Symposium on Edge Computing (SEC)*. IEEE, 2024, pp. 206–219.