

Article

Not peer-reviewed version

U-Net Inspired Transformer Architecture for Multivariate Time-Series Synthesis

[Shyr-Long Jeng](#) *

Posted Date: 3 June 2025

doi: 10.20944/preprints202506.0113.v1

Keywords: attention; CLLC converter; half-bridge; time-series synthesis; electric vehicle



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

U-Net Inspired Transformer Architecture for Multivariate Time-Series Synthesis

Shyr-Long Jeng

Department of Mechanical Engineering, Lunghwa University of Science and Technology,
Taoyuan City 333326, Taiwan; aetsl@gm.lhu.edu.tw; Tel.: +886-2-8209-3211 (ext. 5113)

Highlights

- **Novel Architecture:** Introduces TS-MSDA U-Net, a U-Net inspired transformer architecture, for long-term multivariate time-series synthesis, incorporating multiscale temporal feature extraction and dual attention mechanisms.
- **Broad Applicability & Performance:** Demonstrates strong performance in two distinct applications: achieving high accuracy in real-world EV data synthesis and significantly enhancing signal resolution in advanced power electronic systems.

What are the main findings?

- TS-MSDA U-Net reduces MAE to within $\pm 1\%$ for key EV parameters across 70 real-world trips, outperforming TS-p2pGAN and other baseline U-Net variants.
- Achieves a $36\times$ resolution enhancement in signal reconstruction from low-sampled Analog-to-Digital Converter (ADC) input for a resonant CLLC half-bridge converter, accurately synthesizing complex nonlinear waveforms.

Abstract: This study introduces a Multiscale Dual-Attention U-Net (TS-MSDA U-Net) model, for long-term time-series synthesis. Enhancing the U-Net architecture with multiscale temporal feature extraction and dual attention mechanisms, the model effectively captures complex time-series dynamics. Performance evaluation was conducted across two distinct applications. First, on multivariate datasets collected from 70 real-world electric vehicle (EV) trips, TS-MSDA U-Net achieved mean absolute errors within $\pm 1\%$ for key vehicle parameters, including battery state of charge, voltage, mechanical acceleration, and torque. This represents a substantial two-fold improvement over the baseline TS-p2pGAN model, although the dual attention mechanisms contributed only marginal gains over the basic U-Net. Second, the model was applied to high-resolution signal reconstruction using data sampled from low-speed analog-to-digital converters in a prototype resonant CLLC half-bridge converter. TS-MSDA U-Net successfully captured non-linear synthetic mappings and enhanced the signal resolution by 36 times, while the basic U-Net failed to reconstruct the signals. These findings collectively highlight the potential of U-Net-inspired transformer architectures for high-fidelity multivariate time-series modeling in both real-world EV scenarios and advanced power electronic systems.

Keywords: attention; CLLC converter; half-bridge; time-series synthesis; electric vehicle

1. Introduction

Time-series data, consisting of sequential observations collected at regular intervals, is fundamental to various industrial applications, including electric vehicle (EV) battery monitoring for state of charge (SOC) estimation and high-resolution signal reconstruction in high-frequency resonant CLLC half-bridge converters. Time-series synthesis (TSS) plays a critical role for generating synthetic sequences that accurately replicate the statistical properties and temporal dependencies of real-world data. This technique is vital for augmenting limited datasets, enhancing model generalization while maintaining intrinsic trends and variability [1]. With the increasing prevalence

of time-series data in engineering and cyber-physical systems [2], TSS has become indispensable for building robust models across diverse domains.

Generative adversarial networks (GANs) have demonstrated considerable success in TSS by training a generator to produce synthetic sequences and a discriminator to distinguish them from real data [3]. Models such as TimeGAN [4] excel in capturing complex temporal dependencies, particularly for energy consumption data augmentation. Hybrid GAN-based models further extend these capabilities. For example, CR-GAN [5] integrates convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to jointly extract both spatial and temporal features. Recent advancements include deep convolutional GAN frameworks [6] for enhancing sparse lithium-ion battery datasets while preserving dynamic behavior, and TS-p2pGAN [7], which adapts image-based pixel-to-pixel GAN frameworks to synthesize long-sequence EV data under realistic driving scenarios.

Traditional CNNs are limited by their local receptive fields, restricting their ability to capture long-range dependencies critical for modeling complex temporal patterns [8]. To mitigate this, strategies such as dilated convolutions, pooling-based multiscale feature aggregation, and larger kernel sizes have been proposed. These methods often struggle in capturing diverse multiscale features, leading in suboptimal performance—particularly when segmenting objects or sequences with varying structures. Even atrous convolutions [9], while expanding receptive fields, remain inherently local, limiting global contextual learning.

The U-Net architecture [10], initially designed for medical image segmentation, has demonstrated adaptable for time-series modeling. Its encoder-decoder structure constructs hierarchical representations while preserving critical features through skip connections [11,12]. Despite its strengths, basic U-Net remains constrained in handling highly complex spatial or temporal dynamics, motivating the need for architectural enhancements.

Transformers, first introduced for natural language processing [13], have increasingly been applied to time-series tasks due to their self-attention mechanism, which effectively model long-range dependencies. Attention-based models [14] significantly improve forecasting accuracy by identifying temporal correlations. However, conventional attention modules operate pointwise, often underutilizing broader temporal structures. Hybrid approaches combining CNNs with transformers [15,16] have improved tasks like image super-resolution, where transformer encoders extract global features from CNN maps, and decoders refine details via cross-attention mechanisms.

In medical image segmentation, several U-Net-based architectures with transformer mechanism have emerged, each offering unique advantages and tackling specific computational challenges. TransUNet [17] is a hybrid model that combines a CNN backbone with a vision transformer in its bottleneck stage. This allows TransUNet to leverage the CNN's strength in extracting low and mid-level spatial features while benefiting from the transformer's ability to model long-range dependencies through global self-attention. However, its use of a standard vision transformer can lead to high computational costs due to the non-hierarchical nature of global attention.

Swin-Unet model [18] integrates the Swin Transformer, a hierarchical vision transformer with shifted windows, into the U-Net framework. This design enables efficient local and global context modeling with linear computational complexity. Its multiscale capability is well-suited for varying spatial resolutions. However, its window-based attention may struggle with very large or irregular contexts unless supported by deeper networks.

UNETR [19] adopts a different approach by replacing the convolutional encoder with a vision transformer, while retaining the U-Net decoder and skip connections. This configuration enables early-stage global dependency modeling and achieves high performance in complex segmentation tasks, though at the cost of increased memory and computational demands due to the quadratic complexity of self-attention.

UNETR++ [20] enhances UNETR with efficient paired attention blocks to learn spatial-wise and channel-wise discriminative features for 3D segmentation. Architectural improvements such as deep supervision, multiscale fusion, and refined skip connections improve feature integration and

gradient flow. While UNETR++ often outperforms Swin-Unet and UNETR, it introduces additional complexity, longer training times, and higher hyperparameter sensitivity, limiting its practicality for datasets with limited annotations.

Despite the progress in U-Net variants, many models emphasize spatial attention while underutilizing channel information and inter-feature dependencies. This gap can hinder segmentation quality, especially in tasks requiring rich contextual representation. Addressing both spatial-wise and channel-wise interactions is crucial for more accurate predictions. The U-Net architecture has gained traction in TSS for its ability to capture multiscale features while preserving temporal detail. As research advances, adapting U-Net frameworks for TSS presents promising opportunities for modeling complex temporal structures across multiple scales, offering significant potential for applications in engineering systems and beyond.

The main contributions of this work are summarized as follows:

- **Dual Attention (DA) Mechanisms in U-Net Framework:** The proposed TS-MSDA U-Net model integrates a hierarchical encoder-decoder structure for multiscale temporal feature extraction with DA mechanisms, comprising both sequence attention (SA) and channel attention (CA), effectively capturing complex temporal dynamics in multivariate time-series data.
- **Enhanced TSS for EVs:** The model achieves mean absolute errors (MAEs) within $\pm 1\%$ across key EV parameters (battery SOC, voltage, acceleration, torque) using an open-source dataset from 70 real-world trips, demonstrating a two-fold improvement over baseline TS-p2pGAN model with enhanced alignment to original data distributions.
- **High-Resolution Signal Reconstruction:** The TS-MSDA U-Net achieves a $36\times$ enhancement in signal resolution from low-speed ADC data of a resonant CLLC half-bridge converter, successfully capturing complex nonlinear mappings where the basic U-Net models failed.
- **Cross-Domain Validation and Attention Mechanism Analysis:** The model is validated across two distinct engineering domains: automotive and power electronics, demonstrating robustness and generalizability. While the DA mechanism offers performance gains, analysis reveals its marginal contribution over the basic U-Net in certain cases, providing insights for architectural refinements.

The remainder of this paper is organized as follows. Section 2 introduces the proposed hierarchical encoder-decoder architecture tailored for multiscale temporal learning. It also details DA block, comprising SA and CA modules, which enhance the model's ability to capture complex temporal and inter-feature dependencies in multivariate time-series data. Section 3 presents the experimental results, structured into two parts. The first part focuses on the EV trip dataset, an open-source dataset containing 70 real-world EV driving sessions, and compares the proposed model's performance with the TS-p2pGAN baseline. The second part examines periodic signal reconstruction in a resonant CLLC half-bridge converter. This includes the generation of time-series training data using the PLECS simulator, training evaluation, and testing outcomes from prototype hardware. Section 4 concludes the paper by summarizing the major findings and contributions.

2. Methodology

2.1. Hierarchical Encoder-Decoder Network

The TS-MSDA U-Net architecture, illustrated in Figure 1, is designed for TSS through a hierarchical encoder-decoder network enhanced with dual attention-convolution (DA-conv) blocks. In contrast to the UNETR model [19], which replaces convolutional encoders with transformers, TS-MSDA U-Net preserves the basic U-Net's multilevel structure, facilitating effective multiscale feature extraction.

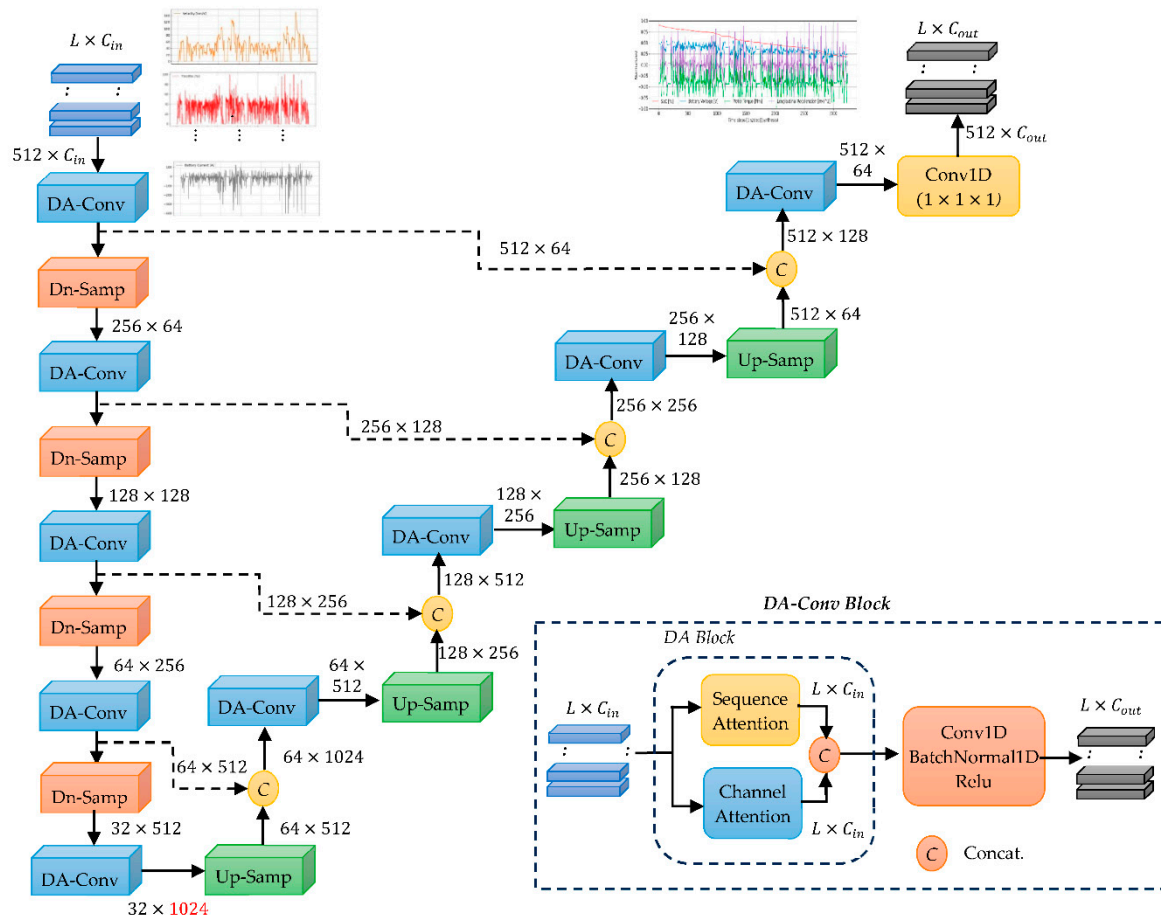


Figure 1. Overview of the TS-MSDA U-Net architecture.

The encoder consists of four stages that progressively downsample the input time-series data. This is accomplished using a DA-conv block followed by one-dimensional max pooling. The input sequence $\mathbf{X} \in \mathcal{R}^{L \times C_{in}}$, where L represents the sequence length and C_{in} the input channel dimension, is configured with $(L=512, C_{in}=10)$ for real-world EV driving data and $(L=512, C_{in}=6)$ for resonant CLLC half-bridge converter data. In the first encoder stage, a DA-conv block projects the channel dimension from C_{in} to 64 channels. Subsequently, each encoder stage applies downsampling through 1D max pooling, halving the sequence length. Simultaneously, DA-conv blocks double the number of feature channels at each stage. The encoding process culminates in a bottleneck layer that applies a final DA-conv block, producing a feature map of size 32×1024 , which is forwarded to the decoder.

The decoder mirrors the encoder's structure with four symmetric stages. Each stage begins with a one-dimensional deconvolution layer that upsamples the feature map, doubling both the temporal resolution and the number of channels. Skip connections connect each encoder stage to its corresponding decoder stage, allowing their outputs to be concatenated and passed through a DA-conv block. At each decoder stage, the number of feature channels is halved to maintain symmetry with the encoder. The final decoder output is merged with earlier convolutional features to restore sequence-level detail and refine representation. A concluding 1×1 convolutional layer then produces the synthesized time-series output with dimensions $512 \times C_{out}$, where C_{out} is the desired number of output channels.

The DA-conv block, a key component of the architecture, integrates a dual attention (DA) module followed by a convolutional block. The DA module captures dependencies along both the sequence and channel dimensions using a key-query attention mechanism. This DA enhances the model's ability to learn joint temporal and inter-channel patterns. The subsequent convolutional block consists of a 1D convolutional layer, batch normalization, and a ReLU activation function. This

combination allows the DA-conv block to efficiently model complex temporal relationships in multivariate time-series data.

A notable strength of TS-MSDA U-Net is the use of high channel dimensionality in the decoder's upsampling path, which facilitates the preservation and propagation of contextual information to higher-resolution layers. The combination of alternating downsampling and upsampling, enriched by skip connections, enables deep hierarchical feature learning and robust time-series reconstruction. The final convolutional layer effectively integrates these features to generate high-fidelity synthetic sequences, recovering information lost during encoding and improving the precision of the output.

2.2. Dual-Attention Block

The dual-attention (DA) block, depicted in Figure 2, is a core component embedded within each DA-conv block across both the encoder (contracting) and decoder (expanding) paths of the TS-MSDA U-Net. This block combines a SA module and a CA module to jointly enhance temporal and inter-channel feature modeling. To ensure the attention mechanisms are aware of positional information, positional embeddings are introduced prior to the attention layers.

2.2.1. Positional Embedding

To provide the model with positional context, each element in the input sequence is assigned a unique positional vector. Given a time-series input $\mathbf{X} \in \mathcal{R}^{L \times C}$, where L is the sequence length and C is the channel dimension, the positional embedding matrix $\mathbf{P} \in \mathcal{R}^{L \times C}$ is constructed. The final embedded input is computed as:

$$\mathbf{X}' = \mathbf{X} + \mathbf{P} \quad (1)$$

where \mathbf{X}' contains both the raw time-series features and positional context, serving as input to the DA modules.

2.2.2. Sequence Attention Module

The SA module aims is designed to model long-range sequence dependencies along temporal axis. It operates on the transposed input feature matrix $(\mathbf{X}')^T \in \mathcal{R}^{C \times L}$, effectively treating each channel as a token sequence. From this transposed input, the query (\mathbf{Q}_s^*), key (\mathbf{K}_s^*), and value (\mathbf{V}_s) matrices are generated via learned linear projections:

$$\mathbf{Q}_s^* = \mathbf{W}^{Q_s} \cdot (\mathbf{X}')^T, \mathbf{K}_s^* = \mathbf{W}^{K_s} \cdot (\mathbf{X}')^T, \text{ and } \mathbf{V}_s = \mathbf{W}^{V_s} \cdot (\mathbf{X}')^T \quad (2)$$

where \mathbf{W}^{Q_s} , \mathbf{W}^{K_s} , and $\mathbf{W}^{V_s} \in \mathcal{R}^{C \times C}$ represent learnable weight matrices. The sequence attention map $\mathbf{A}_s \in \mathcal{R}^{L \times L}$ is computed by measuring the similarity between temporal positions:

$$\mathbf{A}_s = \text{softmax}((\mathbf{K}_s^*)^T \cdot \mathbf{Q}_s^*) \quad (3)$$

This attention map captures dependencies across time steps. The output of the SA module is then calculated as:

$$(\mathbf{O}_s)^T = \mathbf{V}_s \cdot \mathbf{A}_s \quad (4)$$

where $\mathbf{O}_s^T \in \mathcal{R}^{C \times L}$ is subsequently transposed to match the original input shape.

2.2.3. Channel Attention Module

The CA module operates in parallel with the SA module, but focuses on capturing dependencies along the channel dimension. Using the original input $\mathbf{X}' \in \mathcal{R}^{L \times C}$, the module computes the query, key, and value projections:

$$\mathbf{Q}_c^* = \mathbf{W}^{Q_c} \cdot \mathbf{X}', \mathbf{K}_c^* = \mathbf{W}^{K_c} \cdot \mathbf{X}', \text{ and } \mathbf{V}_c = \mathbf{W}^{V_c} \cdot \mathbf{X}' \quad (5)$$

where W^{K_c} , W^{Q_c} , and $W^{V_c} \in \mathcal{R}^{L \times L}$ are learnable parameters. The channel attention map $A_c \in \mathcal{R}^{C \times C}$ is obtained by:

$$A_c = \text{softmax}((K_c^*)^T \cdot Q_c^*) \quad (6)$$

and the output of the CA module is computed as:

$$O_c = V_c \cdot A_c \quad (7)$$

This process allows the CA module to identify salient inter-channel relationships.

2.2.3. Shared Query /Key Projections and Feature Fusion

Given that in most time-series applications the sequence length L significantly exceeds the number of channels C , the computational burden of computing large attention maps becomes a concern. To reduce complexity and promote shared learning, the TS-MSDA U-Net reuses query and key projections from the SA module in the CA module, as follows:

$$Q_c^* = (K_s^*)^T \quad \text{and} \quad K_c^* = (Q_s^*)^T \quad (8)$$

Using this shared representation, the channel attention map can be equivalently expressed as:

$$A_c = \text{softmax}((K_c^*)^T \cdot Q_c^*) = \text{softmax}(Q_s^* \cdot (K_s^*)^T) \quad (9)$$

where $A_c \in \mathcal{R}^{C \times C}$ remains consistent in capturing inter-channel dependencies.

Finally, the outputs of the SA and CA modules are concatenated along the channel dimension to form the final output of the DA block. This fused representation integrates both temporal and channel information, producing a richer and more context-aware feature map for TSS.

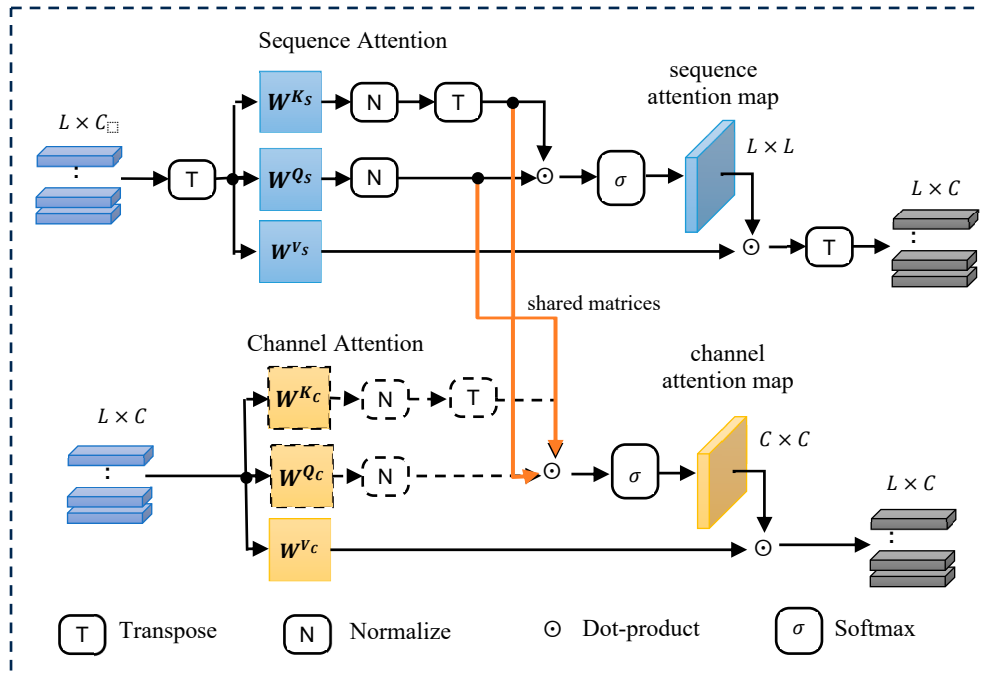


Figure 2. Architecture of DA block.

3. Experimental Setup and Results

To evaluate the effectiveness of U-Net variants in synthesizing long-term time-series data, two distinct experiments were conducted.

The first experiment focused on synthesizing four critical driving features extracted from a dataset comprising 70 EV trips conducted under a range of real-world driving conditions. The performance of the proposed TS-MSDA U-Net was rigorously benchmarked against several baseline and advanced models, including the basic U-Net [10], U-Net with SA, UNETR [19], UNETR++ [20], and previously published results from the TS-p2pGAN model [7]. These comparisons provided a comprehensive assessment of the proposed model's performance in realistic EV scenarios.

The second experiment investigated the enhancement of ADC signal resolution. Interestingly, this experiment revealed that while attention mechanisms offered only marginal improvements over the basic U-Net in relatively simple synthesis tasks, they significantly enhanced the fidelity of generated outputs in more complex, nonlinear scenarios. This highlights the value of attention-based modeling in capturing intricate sequence-wise and channel-wise dependencies.

All models were implemented in Python using the PyTorch deep learning framework. Experiments were executed on a system equipped with an Intel Core i7-10700K CPU and an NVIDIA GeForce RTX 4090 GPU with 24 GB of VRAM. The datasets used in the first and second experiments consisted of 70,090 and 79,872 segments, respectively. For both tasks, the data was randomly partitioned into 80% training and 20% validation sets.

Model training was conducted using the Adam optimizer, with the MAE as the objective loss function. A consistent batch size of 256 was used across all training runs. To ensure training stability and reduce the risk of overfitting, the datasets were fully reshuffled at the beginning of each epoch. Training and validation losses were carefully monitored to assess generalization performance. To quantitatively evaluate the similarity between the real and synthesized time-series, three performance metrics were employed: Root Mean Square Error (RMSE), MAE and Dynamic Time Warping (DTW). The DTW metric [21], was particularly important for capturing temporal misalignments between the real and synthesized sequences. By accounting for non-linear time distortions, DTW offered a more robust and informative assessment of sequence-level similarity compared to point-wise metrics alone.

3.1. Vehicle Trip Dataset

A comprehensive dataset focused on high-voltage battery behavior in EVs was obtained from the IEEE DataPort repository [22]. This dataset comprises real-world driving data collected from 70 distinct trips taken by a BMW i3 (60 Ah) under both summer (Group A) and winter (Group B) conditions. These varied environmental settings provide valuable insights into EV and battery performances under diverse operational and climatic scenarios.

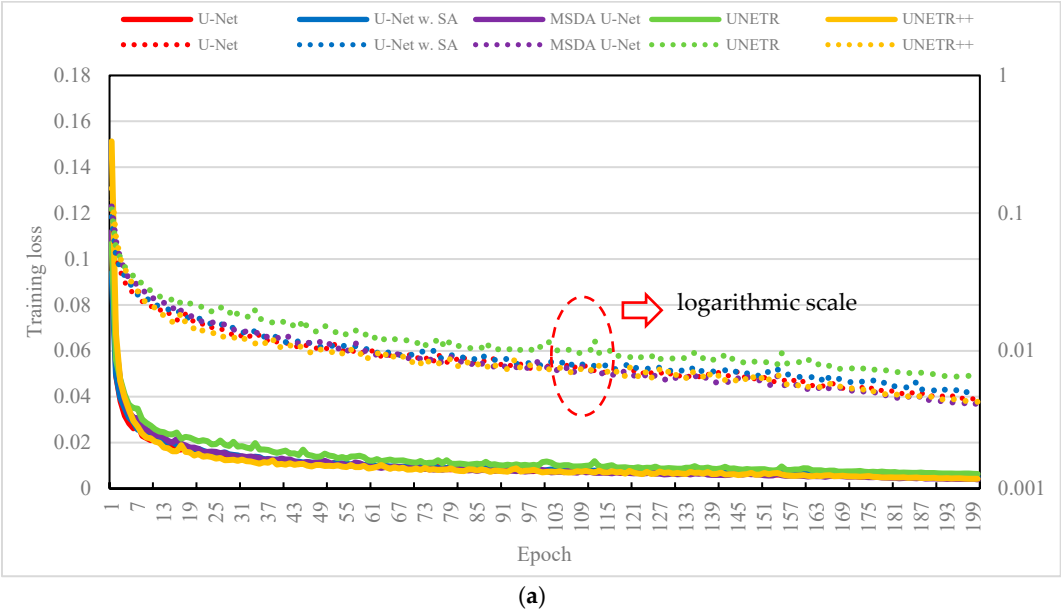
Each trip includes 30 recorded variables, capturing a wide spectrum of information such as environmental conditions, vehicle dynamics, battery performance metrics, and details of the heating system. For training the TS-MSDA U-Net model, 10 key variables were selected based on their relevance and impact. These include vehicle speed, altitude, throttle position, motor torque, acceleration, battery voltage, battery current, battery temperature, State of Charge (SOC), displayed SOC, power consumption of the heater and air conditioner, heater voltage, heater current, and ambient temperature.

The raw dataset, originally sampled at 100-millisecond intervals, was first preprocessed using a 1-second moving average filter, effectively smoothing short-term fluctuations by averaging every ten consecutive samples. The resulting smoothed time-series data was then segmented into overlapping sequences, each containing 512 consecutive data points, using a sliding window approach with a stride of 1. This ensured continuous and comprehensive temporal coverage. To standardize the input and facilitate stable model training, all time-series sequences were normalized to a $[-1, 1]$ range.

3.1.1. Baseline Comparison

Figure 3 presents the training and validation loss curves, measured by MAE, for five model variants. As shown in Figure 3(a), all models demonstrated a steady reduction in training loss, indicating effective convergence. To better highlight subtle differences—particularly during the plateauing stages—a dual y-axis with logarithmic scaling was used. While the baseline U-Net displayed solid overall performance, the TS-MSDA U-Net and UNETR++ achieved slightly lower training losses, indicating enhanced optimization and learning capability. In contrast, U-Net with SA and UNETR showed relatively higher training losses, possibly due to their increased architectural complexity or greater sensitivity to the data's temporal characteristics.

Figure 3(b) shows the corresponding validation loss curves, which generally mirrored the training trends—demonstrating good generalization across models. However, with exception of TS-MSDA U-Net, the other occasionally exhibited spikes in validation loss. These transient fluctuations may be due to the transformer-based models' sensitivity to sequence variability or abrupt changes in input dynamics, or to the U-Net's responses to outlier patterns. Notably, these spikes were often followed by recovery in the next epoch, suggesting a degree of model resilience and adaptation. Despite these fluctuations, the overall model performance remained stable. Importantly, the proposed TS-MSDA U-Net consistently achieved the lowest and most stable validation loss throughout training, underscoring its effective feature learning, and superior predictive accuracy compared to the other model variants.



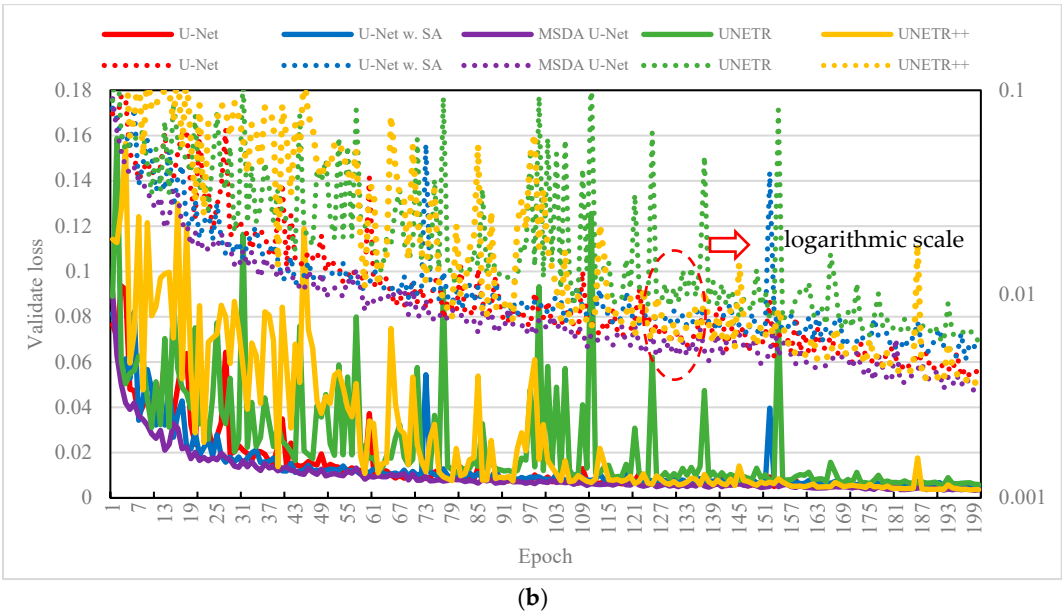


Figure 3. (a) Training and (b) validation loss curves obtained during the training phase.

Figure 4 illustrates the performance of the proposed TS-MSDA U-Net in synthesizing four critical driving parameters—SOC, battery voltage, motor torque, and longitudinal acceleration—throughout the entire 3,240-second duration of the B01 trip. A visual comparison between the synthetic signals (shown in blue) and the corresponding real-world measurements (in red) reveals a high degree of alignment. The green dotted lines, representing the error margins between the synthetic and real signals, remain consistently narrow and well-contained within predefined bounds.

To ensure a fair comparison with the previously reported TS-p2pGAN model [Jeng, 2025], the upper and lower limits of the error bands were set to match the reference ranges used in that study. Two magnified time intervals—specifically [648, 904] and [1942, 2198]—are highlighted to examine the model’s ability to replicate rapid transitions and short-term fluctuations typical of real-world driving conditions. Within these windows, the TS-MSDA U-Net effectively captures high-frequency variations and dynamic changes with minimal lag or distortion.

While minor discrepancies were observed in the synthesized motor torque and longitudinal acceleration, particularly at sharp peaks or during rapid transitions, these deviations were generally small. Crucially, the TS-MSDA U-Net exhibited reduced error magnitudes compared to the TS-p2pGAN, which previously struggled with modeling high-frequency components. The TS-MSDA U-Net maintained error distributions tightly centered around zero, indicating a high degree of temporal fidelity, and generalization capability in replicating complex EV behaviors.

Table I provides a detailed comparative analysis of the five U-Net model variants alongside the previously proposed TS-p2pGAN, evaluated across all driving trips using three key performance metrics: RMSE, MAE, and DTW. Among all the models, the TS-MSDA U-Net consistently outperformed its counterparts, achieving the lowest RMSE, MAE, and DTW values in nearly all driving scenarios. For example, in Trip 1, the TS-MSDA U-Net attained superior performance with an RMSE of 0.68%, an MAE of 0.39%, and a DTW of 0.47%, highlighting its ability to accurately capture the subtle critical dynamics of EV battery behavior during operation.

Compared to the baseline U-Net, neither the U-Net with SA nor the UNETR model demonstrated significant improvements. In fact, UNETR frequently recorded the highest error values among the U-Net variants. A notable case is Trip 41, where UNETR yielded an RMSE of 5.57%, an MAE of 2.94%, and a DTW of 3.34%, the highest observed across all evaluated models. This performance disparity suggests that UNETR’s emphasis on capturing long-range dependencies may be ill-suited for effectively modeling the localized temporal and channel-level variations of EV data. In contrast, the TS-MSDA U-Net is explicitly designed to handle such variations through its multi-scale feature extraction and dual-attention mechanisms.

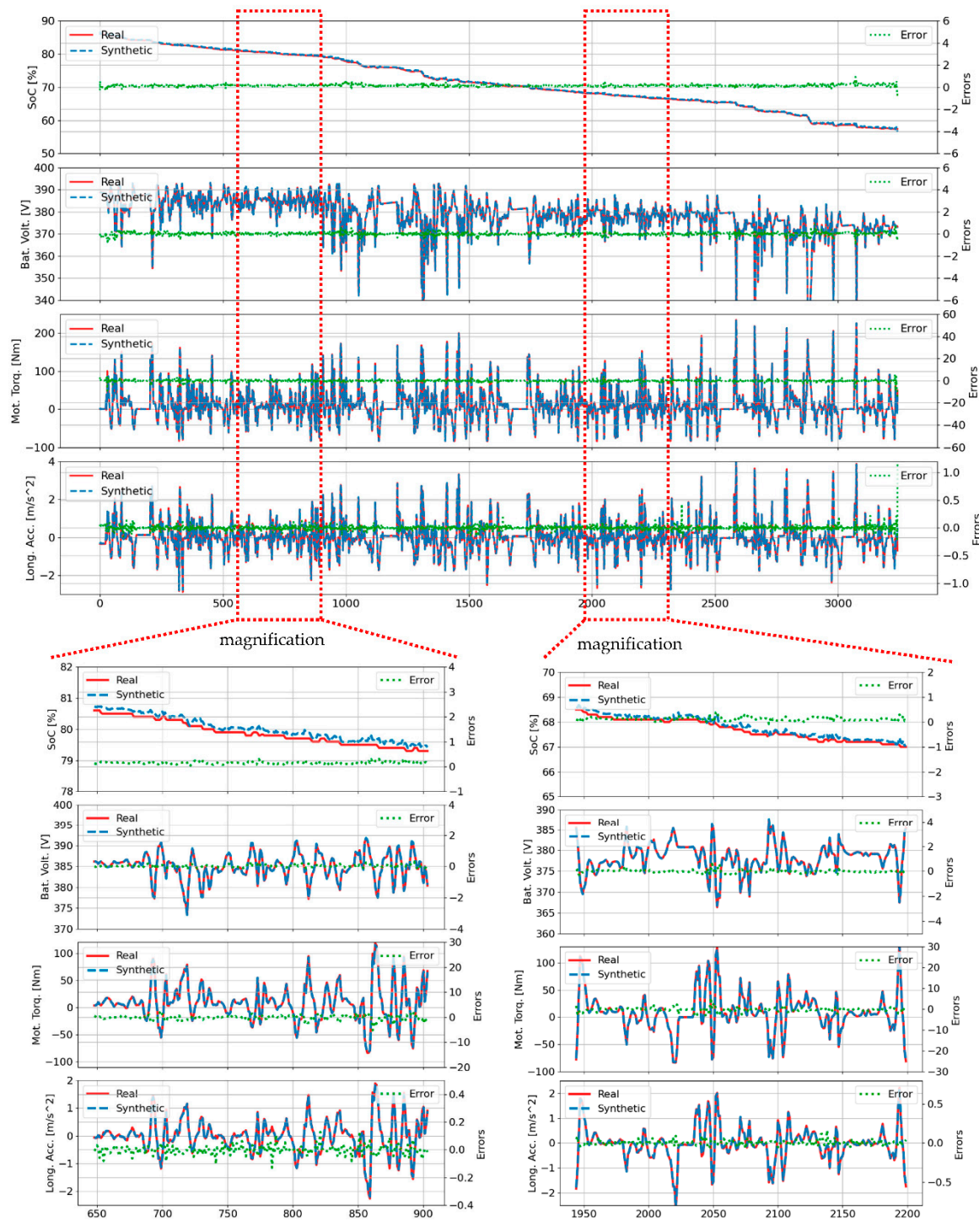


Figure 4. Real and synthetic time-series data generated by the TS-MSDA U-Net model for the B01 trip.

While UNETR++ showed modest improvements over its predecessor, its overall performance still fell short of the TS-MSDA U-Net. Although the two models occasionally produced comparable results—and UNETR++ even matched TS-MSDA U-Net’s performance on a few specific trips—the TS-MSDA U-Net demonstrated greater consistency and reliability across the entire dataset.

In comparison to all U-Net-based models, the TS-p2pGAN exhibited the most inconsistent and generally inferior performance. In numerous driving trips, it recorded substantially higher error metrics, indicating limited generalization capability across diverse driving conditions. For instance,

in Trip B01, the TS-MSDA U-Net achieved an RMSE of 0.68%, an MAE of 0.39%, and a DTW of 0.47%, whereas TS-p2pGAN lagged significantly with values of 1.96%, 0.97%, and 1.19%, respectively.

Overall, the results presented in Table I firmly establish the effectiveness of the proposed TS-MSDA U-Net architecture for TSS in real-world EV driving scenarios. Its ability to incorporate both SA and CA mechanisms enables it to capture the intricate temporal dependencies inherent in EV powertrain dynamics. Moreover, its consistently low error rates, even under highly variable and complex conditions, demonstrate its generalization capability, and practical applicability for real-time signal reconstruction and synthetic data generation in EV systems.

Figure 5 presents a visual comparison of the discrepancies in four key synthesized features between the TS-MSDA U-Net and the previously established TS-p2pGAN models across all driving trips, utilizing violin plots for a comprehensive distribution analysis. The segment length for model input differed between the two approaches: TS-MSDA U-Net was trained using sequences of 512 consecutive samples, while TS-p2pGAN used shorter segments of 256 samples. As a result, Trips 13, 34, and 42 were excluded from the TS-MSDA U-Net evaluation due to their insufficient total number of time steps (i.e., fewer than 512 seconds).

Trained with the MAE loss metric, the TS-MSDA U-Net produced discrepancy distributions with centroids closely centered around zero across all features and driving trips. This indicates a balanced error distribution and strong predictive accuracy. In contrast, TS-p2pGAN exhibited noticeably more variation, with discrepancy distributions showing larger deviations from zero, reflecting less consistent and less reliable predictions under varying conditions.

More specifically, the TS-MSDA U-Net demonstrated tightly concentrated discrepancy ranges within $\pm 1\%$ across all four features, underscoring its high precision and stability across different trips. By comparison, TS-p2pGAN showed greater variability, particularly in motor torque and longitudinal acceleration, where discrepancies occasionally reached up to 5%, indicating a broader error distribution and reduced prediction fidelity.

The shape of the discrepancy distributions further highlighted the TS-MSDA U-Net's consistency. For most trips, the model produced symmetrical, bell-shaped curves, indicative of a well-balanced and normally distributed error profile. A few exceptions were observed, for example, Trips 4, 14, 33, 41, and 62 for SOC, and Trip 41 for battery voltage—which are marked with red boxes in Figure 5(a). In contrast, the TS-p2pGAN model more frequently deviated from this ideal distribution shape, suggesting possible prediction bias or inconsistency. Notable departures from normality were seen in the SOC predictions for Trips 23, 29, 39, 42, 55, and 62, and in battery voltage for Trips 13, 23, 29, 42, 55, and 57—also highlighted with red boxes in the figure.

These findings further reinforce the robustness of the TS-MSDA U-Net model. Its ability to maintain low error magnitudes and consistent distribution shapes across diverse driving scenarios makes it a more accurate and reliable solution for synthesizing key time-series features in electric vehicle applications.

Table 1. RMSE, MAE, and DTW values for all trips.

Trip No	U-Net			U-Net with SA			TS-MSDA-UNet			UNETR			UNETR++			TS-p2pGAN		
	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)
1	0.96	0.45	0.54	1.05	0.52	0.63	0.68	0.39	0.47	1.25	0.61	0.76	0.75	0.37	0.47	1.96	0.97	1.19
2	0.71	0.46	0.53	0.80	0.49	0.57	0.65	0.42	0.51	1.10	0.63	0.69	0.68	0.36	0.46	2.02	1.01	1.12
3	0.72	0.45	0.53	0.83	0.47	0.58	0.59	0.39	0.47	1.17	0.60	0.73	0.65	0.38	0.47	1.91	1.02	1.21
4	1.07	0.42	0.48	1.21	0.47	0.53	0.66	0.35	0.36	1.40	0.51	0.58	0.75	0.35	0.39	1.79	0.77	0.84
5	1.34	0.54	0.65	1.43	0.60	0.74	0.92	0.45	0.55	1.72	0.67	0.83	1.09	0.46	0.59	1.91	0.92	1.09
6	0.81	0.46	0.55	1.01	0.49	0.61	0.65	0.38	0.47	1.11	0.57	0.70	0.78	0.39	0.51	1.62	0.84	1.03
7	0.76	0.42	0.50	0.83	0.44	0.54	0.67	0.39	0.46	1.00	0.54	0.65	0.71	0.38	0.48	1.56	0.84	1.01
8	0.74	0.40	0.47	0.81	0.41	0.50	0.60	0.34	0.41	0.93	0.47	0.56	0.64	0.32	0.40	1.51	0.78	0.92
9	0.77	0.43	0.48	0.86	0.44	0.50	0.54	0.32	0.37	1.08	0.48	0.55	0.62	0.32	0.39	1.73	0.85	0.95
10	1.02	0.51	0.62	1.19	0.53	0.67	0.74	0.41	0.51	1.40	0.63	0.78	0.92	0.42	0.54	2.03	1.06	1.26
11	1.14	0.54	0.61	1.14	0.46	0.55	0.81	0.42	0.49	1.69	0.65	0.74	0.86	0.39	0.47	2.08	1.01	1.15
12	0.72	0.36	0.43	0.78	0.39	0.46	0.56	0.35	0.40	0.96	0.44	0.51	0.62	0.32	0.39	1.26	0.66	0.72
13	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2.69	1.46	1.56
14	1.28	0.44	0.51	1.36	0.45	0.53	1.19	0.39	0.46	1.54	0.55	0.64	1.08	0.34	0.43	1.48	0.80	0.91
15	0.90	0.47	0.56	1.15	0.50	0.60	0.67	0.42	0.50	1.31	0.62	0.76	0.66	0.38	0.49	1.83	0.96	1.17
16	0.73	0.43	0.52	0.82	0.46	0.57	0.62	0.39	0.48	1.04	0.59	0.74	0.69	0.38	0.49	1.92	1.02	1.21
17	0.79	0.48	0.59	0.83	0.44	0.57	0.64	0.41	0.50	1.06	0.58	0.74	0.67	0.39	0.50	2.11	1.07	1.28
18	1.16	0.53	0.63	1.26	0.52	0.63	0.68	0.43	0.52	1.50	0.63	0.73	0.78	0.41	0.52	1.66	0.87	1.01
19	0.84	0.47	0.56	0.97	0.52	0.62	0.66	0.39	0.48	1.24	0.60	0.72	0.71	0.38	0.48	1.80	0.92	1.09
20	0.76	0.43	0.51	0.94	0.49	0.59	0.58	0.37	0.45	1.25	0.62	0.76	0.80	0.38	0.48	1.91	0.96	1.11
21	0.97	0.45	0.55	1.20	0.46	0.57	0.68	0.36	0.44	1.51	0.58	0.71	0.83	0.37	0.47	1.56	0.80	0.97
22	1.00	0.47	0.57	1.15	0.49	0.61	0.68	0.39	0.49	1.25	0.60	0.73	0.82	0.41	0.53	2.05	0.97	1.20
23	1.07	0.54	0.66	1.37	0.59	0.74	0.86	0.46	0.57	1.76	0.73	0.90	1.09	0.49	0.63	2.65	1.31	1.52
24	1.46	0.59	0.73	1.98	0.67	0.88	0.81	0.48	0.60	2.34	0.90	1.14	1.07	0.50	0.66	2.36	1.25	1.42
25	0.78	0.49	0.60	0.80	0.45	0.56	0.60	0.37	0.46	1.03	0.56	0.68	0.66	0.37	0.48	2.23	1.09	1.26
26	1.53	0.64	0.76	1.56	0.64	0.77	1.09	0.46	0.56	2.55	0.90	1.07	1.22	0.46	0.60	2.75	1.30	1.51
27	1.22	0.59	0.70	1.47	0.60	0.72	0.89	0.43	0.52	1.68	0.71	0.86	0.89	0.48	0.58	2.04	1.11	1.30
28	0.96	0.64	0.74	0.98	0.62	0.72	0.70	0.43	0.53	1.30	0.65	0.79	0.75	0.42	0.54	2.21	1.23	1.40
29	0.92	0.59	0.68	1.04	0.61	0.70	0.69	0.45	0.53	1.20	0.67	0.79	0.77	0.51	0.60	2.90	1.67	1.77
30	1.54	0.65	0.76	1.89	0.66	0.77	0.87	0.49	0.58	1.85	0.75	0.89	0.97	0.54	0.64	2.25	1.02	1.22
31	1.15	0.54	0.67	1.29	0.56	0.70	0.64	0.42	0.51	1.54	0.65	0.82	0.74	0.41	0.54	1.86	0.87	1.06
32	1.15	0.56	0.67	1.49	0.60	0.75	0.92	0.51	0.61	1.90	0.73	0.88	0.99	0.48	0.60	2.70	1.30	1.52
33	1.89	0.89	1.04	2.34	0.91	1.08	1.52	0.76	0.91	2.15	0.99	1.17	1.47	0.74	0.91	2.45	1.24	1.60
34	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2.32	1.12	1.32
35	1.02	0.51	0.62	1.15	0.56	0.67	0.71	0.44	0.54	1.46	0.64	0.79	0.79	0.40	0.53	1.95	0.93	1.08
36	0.90	0.43	0.52	1.08	0.45	0.56	0.71	0.37	0.46	1.26	0.55	0.67	0.84	0.36	0.46	1.75	0.86	1.00
37	1.19	0.53	0.62	1.44	0.60	0.71	0.79	0.42	0.52	1.84	0.73	0.90	0.87	0.43	0.53	1.89	0.93	1.13
38	0.94	0.51	0.63	1.14	0.55	0.69	0.70	0.41	0.50	1.52	0.66	0.83	0.81	0.43	0.55	2.30	1.13	1.36
39	1.04	0.53	0.66	1.21	0.57	0.70	0.80	0.41	0.52	1.28	0.65	0.80	0.87	0.44	0.57	2.17	1.18	1.40
40	1.36	0.81	0.95	1.78	1.10	1.35	1.05	0.60	0.72	2.22	1.40	1.66	0.97	0.57	0.71	2.45	1.33	1.67
41	2.42	1.26	1.41	4.71	2.10	2.34	2.81	1.42	1.52	5.57	2.94	3.34	2.45	1.03	1.17	2.40	1.19	1.48

42	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2.71	1.26	1.40
43	0.87	0.45	0.58	1.12	0.58	0.73	0.76	0.39	0.51	1.40	0.68	0.84	0.91	0.45	0.60	2.33	1.22	1.48
44	0.81	0.37	0.47	1.01	0.49	0.59	0.60	0.32	0.40	1.19	0.50	0.63	0.72	0.34	0.45	1.47	0.77	0.90
45	0.60	0.38	0.47	0.74	0.46	0.56	0.53	0.33	0.41	0.86	0.52	0.64	0.61	0.35	0.45	1.58	0.79	0.93
46	0.89	0.43	0.52	1.13	0.58	0.66	0.62	0.32	0.40	1.13	0.54	0.66	0.70	0.37	0.47	2.41	1.20	1.39
47	1.06	0.46	0.57	1.58	0.62	0.78	0.72	0.36	0.45	1.78	0.65	0.84	0.87	0.42	0.56	2.28	1.08	1.30
48	1.21	0.49	0.63	1.40	0.60	0.75	0.90	0.44	0.54	1.57	0.67	0.84	1.05	0.47	0.61	1.72	0.85	1.00
49	0.76	0.40	0.48	0.93	0.49	0.58	0.58	0.34	0.41	0.98	0.51	0.63	0.70	0.36	0.46	2.58	1.24	1.45
50	1.14	0.49	0.62	1.65	0.63	0.80	0.72	0.38	0.47	2.14	0.78	0.98	0.95	0.46	0.60	3.29	1.46	1.75
51	1.24	0.57	0.73	1.60	0.67	0.86	0.92	0.47	0.60	2.47	0.79	1.01	1.12	0.52	0.68	2.19	1.15	1.39
52	1.13	0.55	0.68	1.37	0.68	0.83	0.79	0.44	0.55	1.52	0.73	0.92	0.95	0.52	0.67	2.21	1.09	1.30
53	0.67	0.37	0.47	0.85	0.49	0.60	0.58	0.33	0.42	1.04	0.52	0.66	0.69	0.36	0.47	1.58	0.79	0.94
54	0.86	0.40	0.48	1.05	0.51	0.60	0.55	0.33	0.40	1.39	0.56	0.66	0.84	0.35	0.44	1.45	0.73	0.87
55	0.56	0.38	0.45	0.61	0.41	0.49	0.54	0.35	0.43	0.65	0.44	0.52	0.52	0.31	0.40	2.31	1.14	1.27
56	1.01	0.48	0.58	1.32	0.61	0.73	0.75	0.45	0.55	1.41	0.71	0.84	0.73	0.45	0.56	1.48	0.84	0.99
57	0.99	0.40	0.50	1.22	0.54	0.64	0.74	0.34	0.42	1.45	0.58	0.72	0.72	0.35	0.45	2.54	1.24	1.45
58	0.88	0.40	0.51	1.55	0.54	0.67	0.56	0.32	0.41	1.91	0.60	0.76	0.71	0.38	0.49	1.75	0.94	1.15
59	0.81	0.46	0.57	1.01	0.56	0.69	0.76	0.42	0.53	1.02	0.57	0.71	0.80	0.44	0.57	1.55	0.79	0.97
60	0.99	0.45	0.55	1.24	0.55	0.66	0.76	0.36	0.45	1.54	0.63	0.77	1.14	0.46	0.57	2.05	1.00	1.19
61	0.83	0.39	0.49	1.21	0.54	0.65	0.80	0.36	0.45	1.57	0.57	0.71	0.87	0.37	0.50	2.05	1.03	1.27
62	1.72	0.91	1.04	1.84	1.09	1.24	1.84	0.85	1.01	1.80	0.95	1.12	1.45	0.76	0.94	2.78	1.27	1.43
63	0.88	0.55	0.65	0.99	0.70	0.80	0.65	0.40	0.50	1.07	0.65	0.80	0.78	0.42	0.55	1.87	0.99	1.21
64	0.84	0.47	0.58	0.99	0.56	0.67	0.64	0.37	0.46	1.15	0.62	0.77	0.75	0.40	0.53	2.04	1.06	1.35
65	0.86	0.46	0.58	1.05	0.59	0.71	0.61	0.35	0.45	1.38	0.62	0.80	0.75	0.40	0.53	2.52	1.16	1.45
66	1.28	0.52	0.66	1.44	0.66	0.80	0.74	0.40	0.51	1.79	0.67	0.85	0.93	0.44	0.58	1.63	0.87	1.09
67	0.77	0.44	0.55	0.93	0.51	0.62	0.57	0.34	0.43	1.49	0.58	0.72	0.92	0.39	0.50	2.74	1.37	1.69
68	1.13	0.55	0.67	1.40	0.64	0.79	0.81	0.41	0.53	1.69	0.74	0.93	0.94	0.47	0.62	2.16	1.10	1.35
69	0.81	0.47	0.59	0.98	0.65	0.76	0.65	0.38	0.49	1.09	0.58	0.75	0.79	0.42	0.56	1.88	0.96	1.22
70	1.07	0.51	0.62	1.25	0.60	0.73	0.82	0.38	0.48	2.02	0.73	0.91	0.85	0.41	0.54	2.37	1.13	1.34

Note: The data generated by TS-p2pGAN, sourced from reference [Jeng, 2025], consists of sequences with a length of 256. In this study, the sequence length is extended to 512. Trips 13, 34, and 42 are excluded due to their insufficient duration. Group B trips, labeled B01 to B36, correspond to trip numbers 1 to 36, while Group A trips, labeled A01 to A32, correspond to trip numbers 37 to 70.

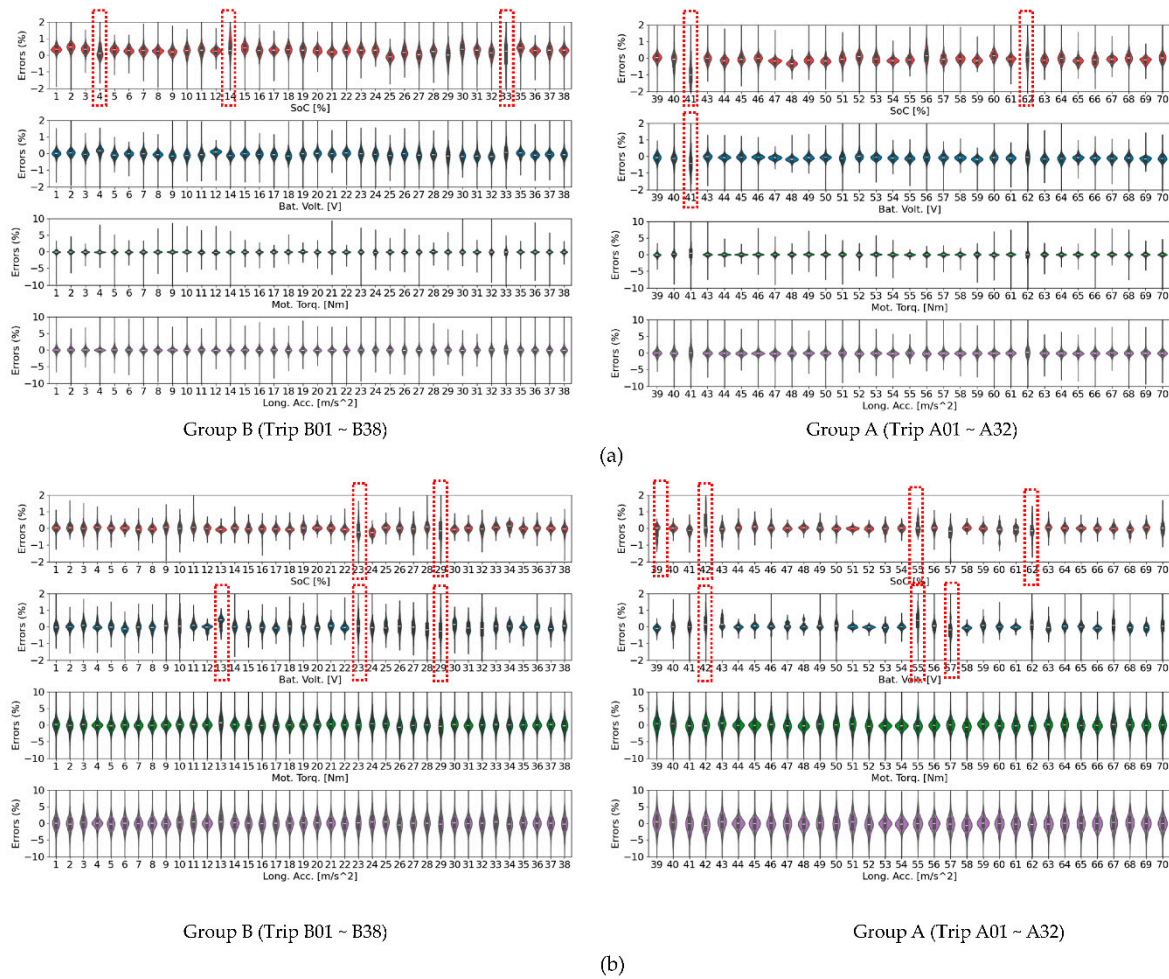


Figure 5. Violin plots comparing error distributions of four features across all trips for: (a) TS-MSDA U-Net; (b) TS-p2pGAN.

3.2. Reconstruction of Periodic Signals for Resonant CLLC Half-bridge Converters

The resonant CLLC half-bridge converter is a highly efficient DC-DC topology widely adopted in applications requiring galvanic isolation and high-frequency operation. These applications include vehicle-to-grid (V2G), vehicle-to-home (V2H) systems, EV fast-charging infrastructure, and e-bike chargers.

As shown in Figure 6, the converter [23] employs an STM32F407VG digital signal processor (DSP) to implement soft-switching control based on both voltage and current feedback. A key architectural highlight is the symmetrical resonant tank, consisting of two inductors (L_1 and L_2) and two capacitors (C_1 and C_2) arranged on either side of a high-frequency transformer. This configuration enables efficient bidirectional power transfer while maintaining excellent voltage regulation across a wide range of operating conditions. The operation of this converter begins with a half-bridge inverter on the primary side, where two Gallium Nitride (GaN) switches (S_1 and S_2) alternate to produce a high-frequency square waveform. When S_1 is turned on and S_2 is off, energy is transferred from the input source through the primary inductor (L_1) and capacitor (C_1) into the transformer's primary winding, initiating resonant energy transfer. During this phase, energy is temporarily stored in the resonant elements and delivered to the load via the transformer and a mirrored CLLC network on the secondary side. As the resonant current oscillates, a brief dead-time occurs—during which both switches are off—allowing the current to naturally reverse direction. This facilitates zero-voltage switching (ZVS) for the next switching cycle, significantly improving conversion efficiency. The converter operates near its resonant frequency, where impedance is minimized and energy transfer is most efficient.

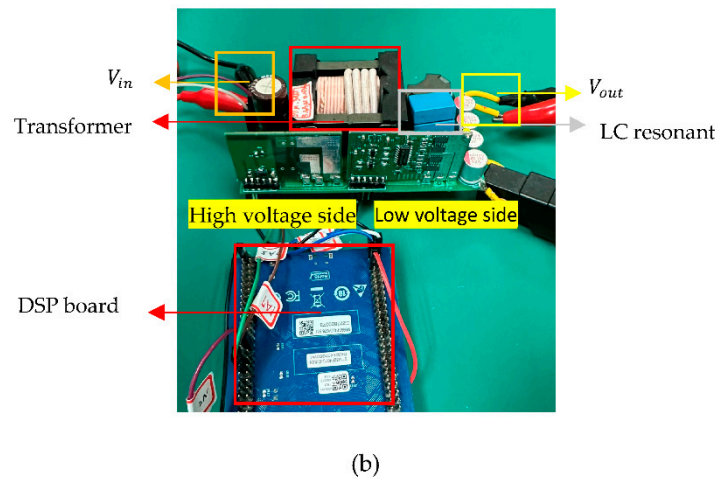
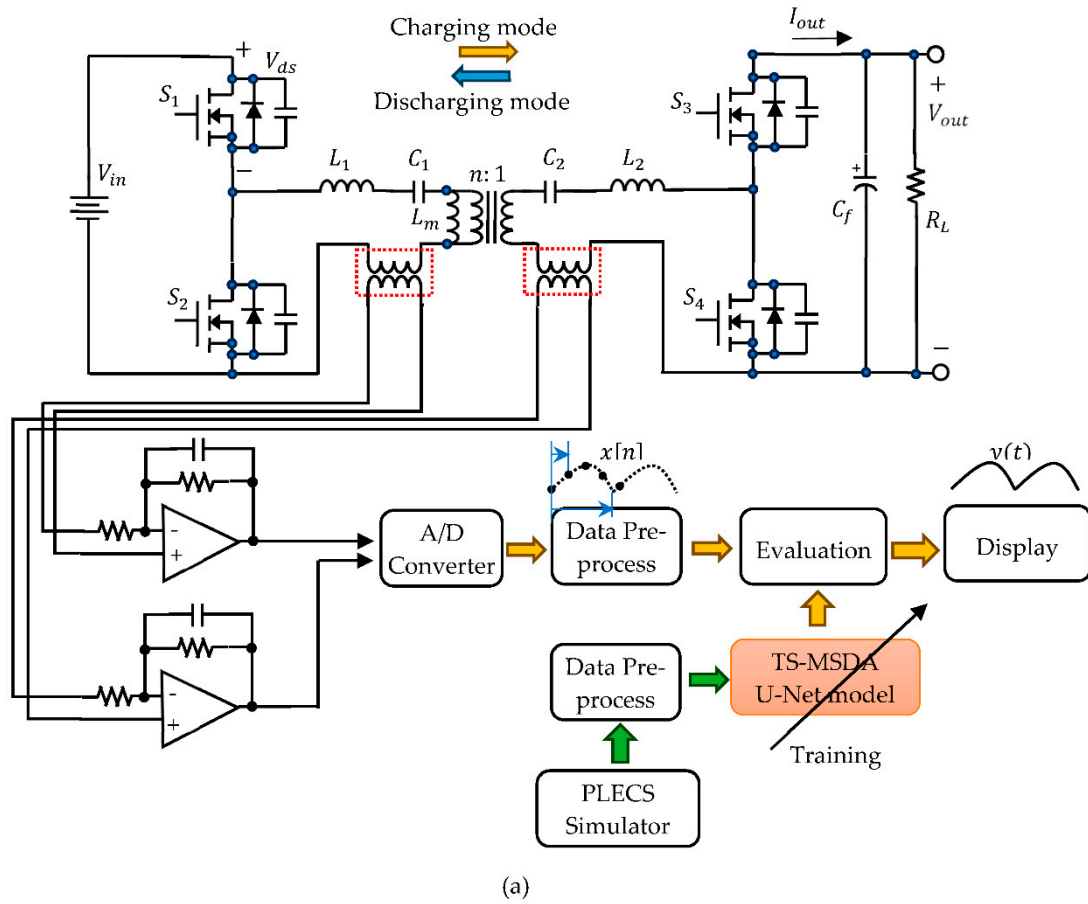


Figure 6. TS-MSDA U-Net architecture applied to enhance A/D resolution in a resonant CLLC half-bridge converter. (a) Circuit schematic. (b) Prototype photograph.

The prototype CLLC converter operates at a switching frequency (f_{pwm}) of 250 kHz. Data acquisition is conducted using an oscilloscope at 500 MS/s via USB, with a resampled rate of 1.25 MHz. Four channels are monitored: the drain-source voltage (V_{ds}) of the upper half-bridge switch on the primary side, the secondary-side capacitor voltage (V_{c2}), the output current (I_{out}) on the secondary side, and the primary-side capacitor voltage (V_{c1}). The primary capacitor voltage (V_{c1}) and the secondary-side capacitor voltage (V_{c2}) serve as critical indicators of the resonance characteristics of the LC tank circuits. However, a key limitation arises due to the insufficient number of ADC samples (only five) collected at 1.25 MHz—too few to capture the full behavior within a single PWM period.

This limitation necessitates the use of signal reconstruction techniques to recover high-resolution periodic signals.

To address this challenge, the coprime sampling technique described in [24] is applied. This method defines two coprime integers, N and M , based on the ratio between the sampling frequency (f_s) and the PWM frequency (f_{pwm}). The integer N represents the number of samples required to reconstruct one full period of the high-resolution signal, while M denotes the number of different PWM cycles across which these samples are collected. The reconstruction process involves arranging these N samples into a full signal period using multiplicative inverse modulo. This algorithm ensures that even with sparse sampling, a high-fidelity periodic signal can be accurately reconstructed.

Crucially, the proposed TS-MSDA U-Net model offers significant advantages over traditional coprime sampling. It enables the DSP to emulate high-resolution signal reconstruction using only its internal ADC, by passing the need for specialized hardware or high-speed external sampling. Moreover, the TS-MSDA U-Net introduces greater flexibility, allowing either the sampling frequency (f_s) or the PWM frequency (f_{pwm}) to be adjusted. This makes it possible to work with non-coprime relationships between N and M , thereby optimizing the resolution and fidelity of the reconstructed signal in various real-world configurations.

3.2.1. Generation of Training Time-Series Data Using the PLECS Simulator

PLECS (Piecewise Linear Electrical Circuit Simulation) [25] is a powerful simulation platform specifically tailored for modeling power electronic systems. Its ability to accurately replicate the switching behavior of semiconductor devices makes it particularly well-suited for simulating resonant converters such as the resonant CLLC half-bridge topology.

To obtain the time-series data required for training the TS-MSDA U-Net model, a detailed simulation of the resonant CLLC half-bridge converter is constructed within the PLECS environment. This converter incorporates appropriately positioned measurement blocks to capture key electrical signals, including V_{ds} , V_{C2} , I_{OUT} and V_{C1} . Time-series data is extracted using the scope block's export function available in PLECS. Once the simulation is executed across various load conditions and over a defined time interval, the generated data is saved as a CSV file for downstream processing.

Prior to model training, the raw simulation data undergoes a preprocessing phase. This includes normalization to ensure consistent data ranges, and segmentation of the continuous waveforms into paired input-output sequences of 512 consecutive time steps, specifically tailored for the A/D resolution enhancement task. This structured data preparation ensures compatibility with the TS-MSDA U-Net training pipeline and enables effective learning of the underlying temporal patterns.

Ultimately, the simulation results from PLECS serve a dual purpose: they not only provide high-fidelity training data for the proposed model but also help validate the functional performance of the resonant CLLC half-bridge converter across diverse operating scenarios. These insights can later be corroborated through experimental validation using a physical prototype of the converter.

3.2.2. Analysis of Training Experimental Results.

Figure 7 illustrates a comparative analysis of the training and validation loss curves for five models developed to enhance the resolution of periodic signals in the resonant CLLC half-bridge converter. The plots use twin y-axes: the right axis features a logarithmic scale, allowing for a clearer representation of loss variations across a wide dynamic range.

Among the models, UNETR (green line) displays a unique loss pattern during the initial training phase. It experiences a sharp decline in training loss during the first few epochs, followed by a temporary increase before resuming a downward trend. This behavior suggests that the model rapidly captures basic data features but initially overfits to specific patterns, leading to transient instability before settling into more generalized learning.

A closer examination of the curves reveals that the baseline U-Net (red line) and UNETR (green line) maintain relatively high and stable loss values throughout training and validation. This

indicates a limited ability to effectively learn and generalize from the training data. In contrast, the other models exhibit substantial performance improvements, with both training and validation losses consistently decreasing overtime. In particular, the TS-MSDA U-Net (purple line) and UNETR++ (yellow line) achieve the lowest losses in both training and validation, especially in the later epochs where their curves converge closely. The application of a logarithmic scale on the secondary axis further highlights their superior performance at finer resolution scales.

During validation, the TS-MSDA U-Net, UNETR++, and U-Net with SA models occasionally show transient spikes in validation loss. However, these spikes are typically followed by subsequent reductions in validation MAE, indicating that these models can recover from temporary overfitting and continuing to refine their predictions. Overall, the TS-MSDA U-Net consistently outperforms the other models, achieving the lowest overall loss and delivering the most effective resolution enhancement for periodic signals in power electronic systems.

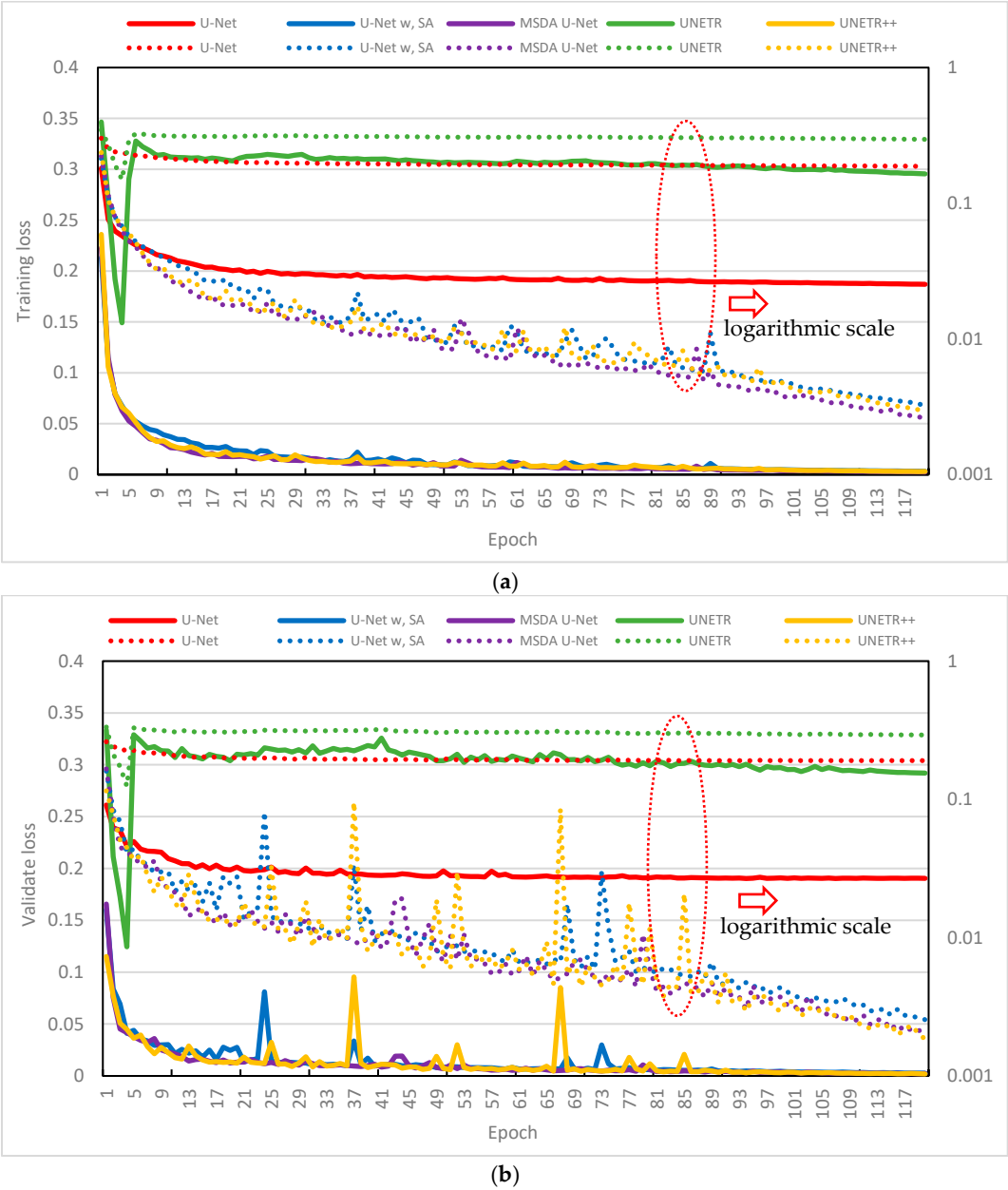


Figure 7. (a) Training and (b) validation loss curves for enhancing the resolution of periodic signals in resonant CLLC half-bridge converters.

Figure 8 illustrates the TS-MSDA U-Net model’s effectiveness in enhancing periodic signals under both heavy and light load conditions. The model synthesizes high-resolution signals from low-

resolution input data generated through PLECS simulations, increasing the sampling density from 5 points per PWM period to 30 and 180 points per PWM period. These inputs are based on a PWM switching frequency of 250 kHz and a simulated ADC sampling frequency of 1.25 MHz. The leftmost panels of Figure 8 show the original low-resolution signals for four critical input channels: V_{ds} , V_{C2} , I_{OUT} and V_{C1} . Each signal is plotted as a solid-colored line, with legend labels indicating the signal's vertical scale (e.g., " V_{ds} (100V)" denotes a grid scale of 100 V per vertical grid). The middle and rightmost panels display the corresponding synthetic signals generated by the TS-MSDA U-Net, overlaid with the ground-truth data. In these enhanced panels, the model's outputs are shown as solid-colored lines, while the ground-truth waveforms are rendered as black dotted lines for direct visual comparison.

The PLECS simulations were conducted under 13 distinct loading conditions, with output current values ranging from 31 A (heavy load) to 1 A (light load). Under heavy-load conditions, the synthetic signals reveal pronounced oscillations and dynamic behaviors across all channels. The output current (I_{OUT}) reaches approximately 31 A, and the resonant capacitor voltage (V_{C1}) swings dramatically within a range of [-200 V, 300 V]. In contrast, under light-load conditions, (V_{C1}) is significantly reduced, fluctuating within a narrower range of [29 V, 72 V], and the waveforms exhibit more stable, less oscillatory behavior, with (I_{OUT}) near 1 A.

Despite the broad variation in operating conditions, the TS-MSDA U-Net demonstrates accuracy in reconstructing both coarse features and fine-grained high-frequency components of the signals. The model successfully recovers high-resolution characteristics from sparsely sampled inputs, underscoring its effectiveness in enhancing temporal resolution for precise signal analysis in the resonant CLLC half-bridge converters.

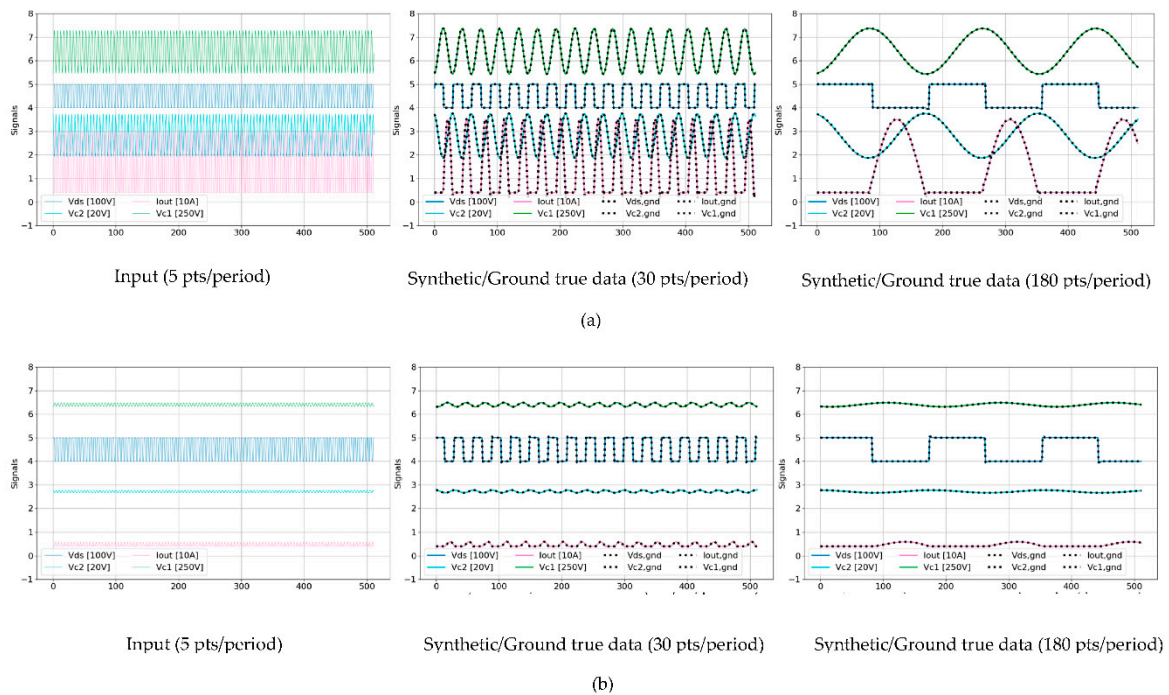


Figure 8. Enhancement of 512-point segments using the TS-MSDA U-Net. The model increases the sampling density from a low resolution of 5 points/period to 30 points/period and 180 points/period under (a) heavy load and (b) light load operating conditions.

Figure 9 presents a comparative visualization of synthetic outputs produced by four distinct models: (a) U-Net, (b) UNETR, (c) U-Net with SA, and (d) UNETR++, evaluated against their corresponding ground truth signals. The results from models (a) and (b) clearly indicate limited reconstruction capability. Both the basic U-Net and UNETR models exhibit notable discrepancies and

deviations in their synthetic outputs when compared to the ground true waveforms, reflecting an insufficient ability to capture the complex temporal dynamics inherent in the data. In contrast, the models enhanced with attention mechanisms—namely (c) U-Net with SA and (d) UNETR++—demonstrate markedly superior performance. Their reconstructed waveforms closely align with the ground truth signals, indicating that the incorporation of attention modules significantly improves the model's ability to learn and replicate the underlying features of periodic signal data. This performance gain underscores the critical role of attention-based architectures in enhancing the fidelity of waveform reconstruction.

Furthermore, when compared to the TS-MSDA U-Net results presented earlier in Figure 8, both the U-Net with SA and UNETR++ models exhibit comparable, and in some instances potentially superior, reconstruction accuracy. Their ability to synthesize signals with high alignment to the ground truth highlights the effectiveness of attention mechanisms in modeling complex signal behavior. As a result, these models emerge as strong candidates for high-resolution time-series reconstruction tasks in the applications requiring precise signal reconstruction under varying operating conditions.

A comprehensive evaluation of the five models—based on RMSE, MAE, and DTW metrics as summarized in Table 2—highlights distinct differences in their performance when enhancing the resolution of periodic signals. These signals were obtained from simulations of a resonant CLLC half-bridge converter operating under thirteen diverse load conditions, ranging from heavy (Case 1) to light (Case 13).

Among the evaluated models, the basic U-Net and UNETR consistently exhibit the highest error values across all load scenarios, demonstrating limited ability to reconstruct accurate signals. Their elevated RMSE, MAE, and DTW values suggest insufficient capacity to capture the complex dynamics present in the time-series data, thus resulting in lower fidelity reconstructions. In contrast, the models that incorporate attention mechanisms—namely, U-Net with SA, TS-MSDA U-Net, and UNETR++—show substantial improvements in reconstruction accuracy. These models achieve significantly lower error values across all metrics and load conditions. In particular, the TS-MSDA U-Net outperforms all others, consistently achieving the lowest RMSE, MAE, and DTW scores in nearly all scenarios. Its advantage is especially prominent under light load conditions, such as Case 13, where accurate reconstruction is more challenging due to the reduced signal amplitude and diminished dynamic range.

The TS-MSDA U-Net's superior performance can be attributed to its dual-attention mechanisms and multiscale feature extraction, which enable the model to effectively learn complex temporal dependencies and capture signal features across varying resolutions. While UNETR++ and U-Net with SA demonstrate better accuracy than its baseline variants, it generally falls short of the TS-MSDA U-Net's performance. Nevertheless, it still achieves a marked improvement over the traditional approach.

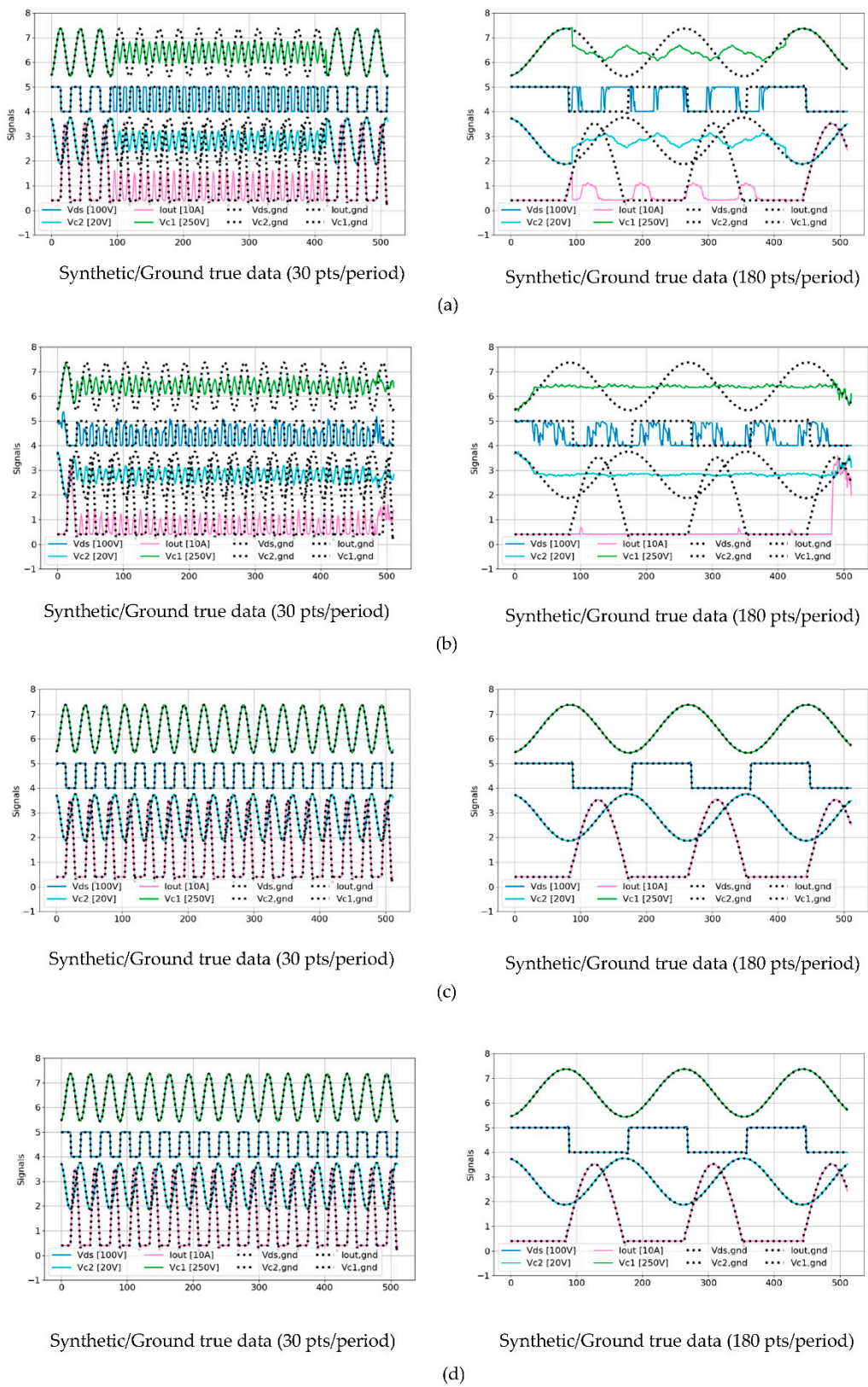


Figure 9. Visual comparisons of synthetic outputs against ground truth data across various models: (a) U-Net, (b) UNETR, (c) U-Net with SA, and (d) UNETR++.

Table 2. RMSE, MAE, and DTW metrics for evaluating resolution enhancement in bidirectional half-bridge CLLC converters.

Case No	U-Net			U-Net with SA			TS-MSDA-UNet			UNETR			UNETR++		
	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)	RMSE (%)	MAE (%)	DTW (%)
1	52.60	30.19	20.17	1.34	0.44	0.47	0.84	0.40	0.41	63.37	45.57	31.28	1.05	0.43	0.47
2	49.11	28.04	18.80	1.15	0.31	0.34	0.90	0.33	0.35	59.09	42.35	29.26	0.89	0.37	0.37
3	46.20	26.28	17.59	1.36	0.29	0.33	0.82	0.30	0.32	55.48	39.57	27.30	1.04	0.33	0.33
4	43.30	24.49	16.09	1.41	0.29	0.32	0.67	0.26	0.28	52.14	37.12	25.38	0.65	0.28	0.29
5	39.79	22.32	14.85	1.35	0.26	0.28	0.65	0.24	0.25	48.45	34.49	23.23	0.48	0.24	0.24
6	37.20	20.60	13.65	1.24	0.24	0.27	0.81	0.22	0.24	45.25	31.98	21.45	0.69	0.22	0.22
7	34.74	18.90	12.41	1.30	0.24	0.27	0.72	0.20	0.22	42.21	29.49	19.63	0.74	0.20	0.20
8	31.68	16.60	10.70	1.36	0.25	0.29	0.53	0.18	0.19	38.54	26.28	17.18	0.51	0.18	0.18
9	29.62	14.86	9.40	1.48	0.23	0.27	0.54	0.17	0.18	35.65	23.49	15.33	0.63	0.17	0.17
10	27.76	13.10	8.11	1.07	0.22	0.25	0.55	0.16	0.18	33.62	21.14	13.49	0.68	0.18	0.18
11	26.24	11.34	6.85	0.51	0.18	0.19	0.30	0.14	0.14	32.09	18.74	11.65	0.28	0.16	0.15
12	25.12	9.56	5.56	0.39	0.16	0.16	0.27	0.12	0.12	30.92	16.18	9.80	0.27	0.15	0.14
13	24.47	7.89	4.32	0.41	0.13	0.14	0.25	0.10	0.10	30.16	13.62	8.09	0.28	0.16	0.15

Note: Case 1 represents a heavy load condition, while Case 13 corresponds to a light load condition. The sequence of cases transitions progressively from heavy load to light load.

3.2.3. Testing Experimental Results Using the Prototype Converters

Figure 10 presents a visual comparison between the synthetic signals generated by the TS-MSDA U-Net and the corresponding ground truth measurements obtained from a prototype CLLC half-bridge converter operating under both heavy and light load conditions. The ground truth signals were captured using an oscilloscope with a sampling rate of 500 MS/s via USB, as shown in the upper-left panel of the figure. To ensure all relevant signals (V_{ds} , V_{C1} , and V_{C2}) were clearly displayed, the oscilloscope grid scales were adjusted according to the load conditions.

Under heavy-load operation, the synthetic data generated by the TS-MSDA U-Net exhibits a high degree of fidelity, with waveform shapes that closely match those of the real measured signals. A minor discrepancy is observed in the V_{ds} waveform, where the synthetic output displays sharper transitions in the square waveforms compared to the more rounded edges of the measured signals—likely due to parasitic effects not captured in the simulation-based training data. Under light-load conditions, the differences between the synthetic and ground truth signals become more pronounced. The overall accuracy of waveform reconstruction is reduced, with a noticeable phase shift appearing in the V_{ds} signal, indicating a slight temporal misalignment between the generated and actual data.

The input provided to the TS-MSDA U-Net was at a low resolution of only 5 points per PWM period, corresponding to an ADC sampling rate of 1.25 MHz. Despite this sparse input, the model successfully reconstructs high-resolution waveforms with 30 and 180 points per period, equivalent to effective sampling rates of 7.5 MHz and 45 MHz, respectively. These results underscore the TS-MSDA U-Net's capability to significantly enhance temporal resolution, even under real-world operating conditions.

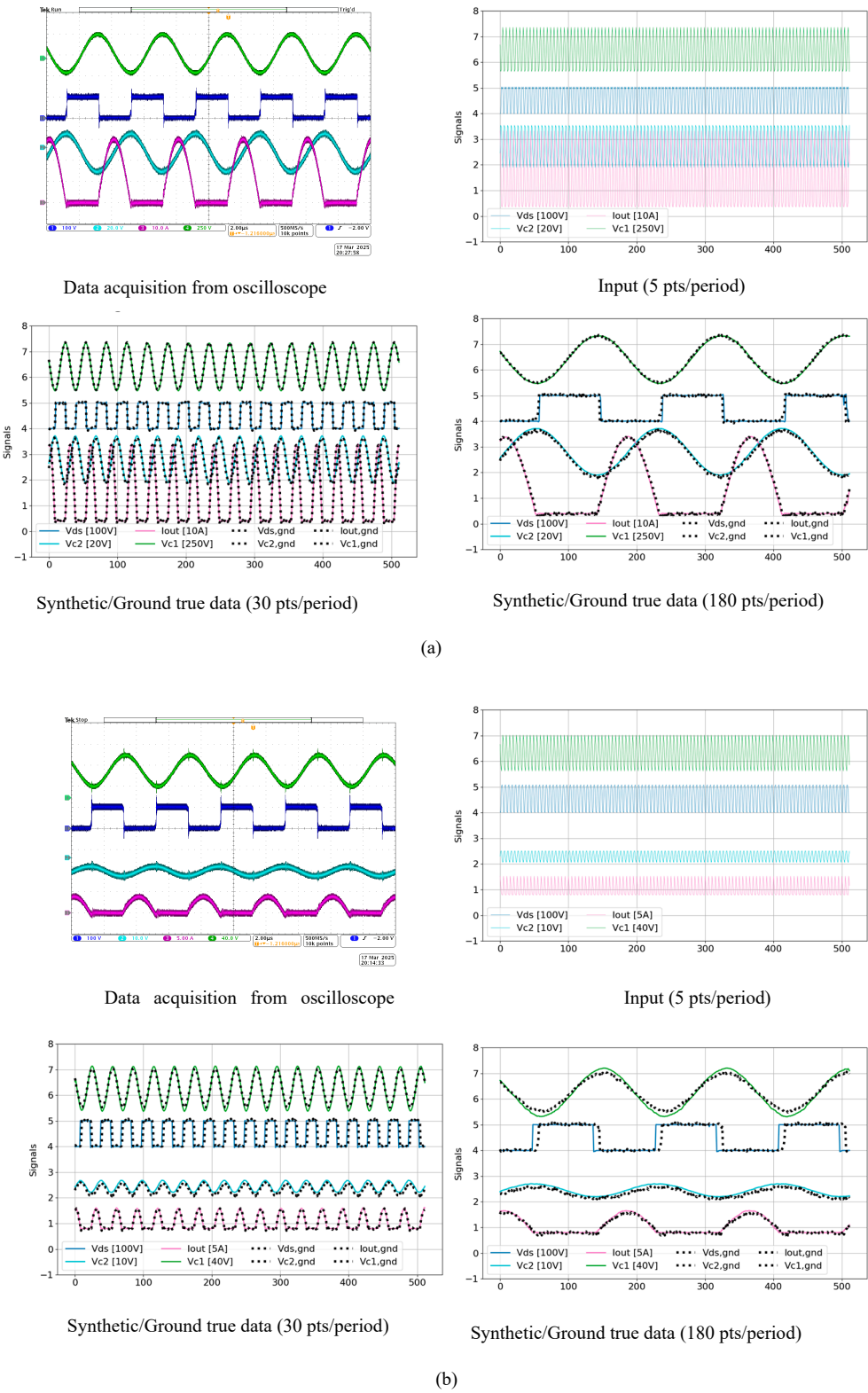


Figure 10. Visual assessment of TS-MSDA U-Net synthetic data against ground truth in a prototype CLLC converter under various load conditions (a) heavy-load and (b) light-load.

4. Conclusion

The TS-MSDA U-Net is a novel architecture specifically designed to capture complex temporal dynamics in long-range TSS tasks. It combines a hierarchical encoder-decoder structure for multiscale temporal feature extraction with a dual-attention mechanism, incorporating both SA and CA to enhance modeling capacity. The performances of the mode were rigorously benchmarked against several baseline architectures, including the basic U-Net, U-Net with SA, UNETR, and UNETR++, using RMSE, MAE, and DTW as key evaluation metrics.

In the first application—synthesizing multivariate time-series data from 70 real-world EV driving trips—the TS-MSDA U-Net demonstrated notable improvements. It achieved MAEs within $\pm 1\%$ for key parameters such as SOC, battery voltage, mechanical acceleration, and motor torque, representing a two-fold improvement over the TS-p2pGAN baseline. While violin plot analyses revealed greater variability in TS-p2pGAN, particularly with deviations up to 5% in motor torque and longitudinal acceleration, the TS-MSDA U-Net showed significantly more stable accuracy. The U-Net with SA and UNETR models exhibited comparatively higher training losses, likely due to their sensitivity to temporal fluctuations or increased architectural complexity. Although UNETR++ occasionally matched the TS-MSDA U-Net in some metrics, the proposed model consistently outperformed the basic U-Net. It is worth noting, however, that in this relatively straightforward EV synthesis task, the DA mechanism provided only marginal gains over the basic U-Net. This observation underscores a broader insight: attention mechanisms are not universally advantageous, and their effectiveness is highly dependent on the data characteristics and task complexity.

The benefits of the TS-MSDA U-Net became far more evident in the second application—high-resolution signal reconstruction for a resonant CLLC half-bridge converter. Trained on simulation data from PLECS and evaluated oscilloscope measurements from a physical prototype, the model demonstrated a 36-fold increase in effective resolution, reconstructing low-resolution ADC input (sampled at 1.25 MHz) into waveforms equivalent to 45 MHz. This capability enabled accurate recovery of nonlinear periodic signal mappings across a wide range of load conditions. In this more demanding task, the DA mechanism played a crucial role in enhancing synthesis fidelity. In contrast, both the basic U-Net and UNETR architectures failed to generate reliable high-resolution reconstructions.

In summary, the TS-MSDA U-Net exhibits robust potential for high-fidelity multivariate time-series modeling, especially in domains characterized by nonlinear dynamics and intricate temporal patterns. Its strong performance in both real-world EV signal synthesis and high-resolution power electronics applications demonstrates the value of integrating attention mechanisms and multiscale feature representations into U-Net architectures for advanced TSS tasks.

Funding: This research was funded by the National Science Council of Taiwan, R.O.C., under Contract NSTC 113-2622-E-262-001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The author declares that they have no conflicts of interest.

References

1. Iglesias, G.; Talavera, E.; González-Prieto, Á.; Mozo, A.; Gómez-Canaval, S. Data Augmentation Techniques in Time Series Domain: A Survey and Taxonomy. *Neural Comput. Appl.* **2023**, *35*, 10123–10145.
2. Sommers, A.; Ramezani, S.B.; Cummins, L.; Mittal, S.; Rahimi, S.; Seale, M.; Jaboure, J. Generating Synthetic Time Series Data for Cyber-Physical Systems. In *Proceedings of the 2024 IEEE 10th World Forum on Internet of Things (WF-IoT)*, Boston, MA, USA, 1–7 November 2024; pp. 1–7.

3. Brophy, E.; Wang, Z.; She, Q.; Ward, T. Generative Adversarial Networks in Time Series: A Systematic Literature Review. *ACM Comput. Surv.* **2023**, *55*, 1–31.
4. Tang, P.; Li, Z.; Wang, X.; Liu, X.; Mou, P. Time Series Data Augmentation for Energy Consumption Data Based on Improved TimeGAN. *Sensors* **2025**, *25*, 493.
5. Tian, Y.; Peng, X.; Zhao, L.; Zhang, S.; Metaxas, D.N. CR-GAN: Learning Complete Representations for Multi-View Generation. arXiv 2018, arXiv:1806.11191.
6. El Fallah, S.; Kharbach, J.; Hammouch, Z.; Rezzouk, A.; Jamil, M.O. State of Charge Estimation of an Electric Vehicle's Battery Using Deep Neural Networks: Simulation and Experimental Results. *J. Energy Storage* **2023**, *62*, 106904.
7. Jeng, S.L. Generative Adversarial Network for Synthesizing Multivariate Time-Series Data in Electric Vehicle Driving Scenarios. *Sensors* **2025**, *25*, 749.
8. Rotem, Y.; Shimoni, N.; Rokach, L.; Shapira, B. Transfer Learning for Time Series Classification Using Synthetic Data Generation. In *International Symposium on Cyber Security, Cryptology, and Machine Learning*; Springer: Cham, Switzerland, 2022; pp. 232–246.
9. Lian, X.; Pang, Y.; Han, J.; Pan, J. Cascaded Hierarchical Atrous Spatial Pyramid Pooling Module for Semantic Segmentation. *Pattern Recognit.* **2021**, *110*, 107622.
10. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Cham, Switzerland, 2015; Volume 9351.
11. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057.
12. Azad, R.; Aghdam, E.K.; Rauland, A.; Jia, Y.; Avval, A.H.; Bozorgpour, A.; Merhof, D. Medical Image Segmentation Review: The Success of U-Net. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, doi:10.1109/TPAMI.2024.3376785.
13. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
14. Xu, L.; Qin, J.; Sun, D.; Liao, Y.; Zheng, J. Pfformer: A Time-Series Forecasting Model for Short-Term Precipitation Forecasting. *IEEE Access* **2024**, doi:10.1109/ACCESS.2024.3387860.
15. Zhu, H.; Xie, C.; Fei, Y.; Tao, H. Attention Mechanisms in CNN-Based Single Image Super-Resolution: A Brief Review and a New Perspective. *Electronics* **2021**, *10*, 1187.
16. Wan, A.; Chang, Q.; Khalil, A.B.; He, J. Short-Term Power Load Forecasting for Combined Heat and Power Using CNN-LSTM Enhanced by Attention Mechanism. *Energy* **2023**, *282*, 128274.
17. Chen, J.; Mei, J.; Li, X.; Lu, Y.; Yu, Q.; Wei, Q.; Zhou, Y. TransUNet: Rethinking the U-Net Architecture Design for Medical Image Segmentation Through the Lens of Transformers. *Med. Image Anal.* **2024**, *97*, 103280.
18. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. In *European Conference on Computer Vision*; Springer Nature Switzerland: Cham, Switzerland, 2022; pp. 205–218.
19. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Xu, D. Unetr: Transformers for 3D Medical Image Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 4–8 January 2022; pp. 574–584.
20. Shaker, A.; Maaz, M.; Rasheed, H.; Khan, S.; Yang, M.H.; Khan, F.S. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2024**, *43*, 3377–3390.
21. Senin, P. Dynamic Time Warping Algorithm Review. *Inf. Comput. Sci. Dep. Univ. Hawaii Manoa Honolulu USA* **2008**, *855*, 40.
22. Battery and Heating Data for Real Driving Cycles. IEEE DataPort 2022. Available online: <https://iee-dataport.org/open-access/battery-and-heating-data-real-driving-cycles> (accessed on 17 November 2024).
23. Shieh, Y.T.; Wu, C.C.; Jeng, S.L.; Liu, C.Y.; Hsieh, S.Y.; Haung, C.C.; Chang, E.Y. A Turn-Ratio-Changing Half-Bridge CLLC DC–DC Bidirectional Battery Charger Using a GaN HEMT. *Energies* **2023**, *16*, 5928.

24. Tang, H.C.; Shieh, Y.T.; Roy, R.; Jeng, S.L.; Chieng, W.H. Coprime Reconstruction of Super-Nyquist Periodic Signal and Sampling Moiré Effect. *IEEE Trans. Ind. Electron.* **2025**, doi:10.1109/TIE.2024.3475960.
25. Asadi, F.; Eguchi, K. *Simulation of Power Electronics Converters Using PLECS®*; Academic Press: Cambridge, MA, USA, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.