

RGANFormer: Relativistic Generative Adversarial Transformer for Time-Series Signal Forecasting on Intelligent Vehicles

Yang Xing* *Senior Member, IEEE*, Xiangqi Kong, Antonios Tsourdos

Abstract—Time-series modelling (TSM) is a critical task for intelligent vehicles (IVs), covering areas like fault detection, health monitoring, and inference of road user intentions. In this study, we present a novel TSM approach for enhancing the accuracy of multi-variate signal forecasting in intelligent vehicles. Our method leverages advanced Transformer networks within a relativistic generative adversarial network (RGAN) training framework. The RGAN training framework efficiently improves the accuracy of vehicle states forecasting for IV, demonstrating effective learning of long-time dependencies for more accurate predictions over extended sequences. Additionally, we introduce a high-dimensional extension (HDE) built-in block for the time-series Transformer to explore the impact of higher-dimensional features on representing long-term sequences. The experimental data is collected from a real-world electric vehicle testing bed. We evaluate the proposed RGANFormer framework and the HDE block on two popular time-series models, namely, Autoformer and FiLM. The results demonstrate that the RGANFormer, along with the built-in HDE block, significantly enhances long-term sequential forecasting accuracy for both multivariate and univariate tasks.

Index Terms—Time-series, generative adversarial network, Transformer, multi-task learning

I. INTRODUCTION

A. Motivation

Time series modelling (TSM) finds extensive applications in intelligent vehicles (IVs). For instance, TSM-based CAN bus signal estimation plays a crucial role in vehicle fault detection and tolerance [1], cyber-security [2], and energy optimization [3]. Accurate forecasting of the vehicle states can significantly mitigate risks from system faults, malfunctions, and cyber-attacks. Traditional TSM methods for IVs often rely on conventional signal processing and machine learning techniques such as support vector machines [4], random forests [5], and feedforward neural networks [6]. With the recent advancements in deep learning, including recurrent neural networks [7], and Transformer networks [8], [9], TSM has become increasingly powerful for sequence modelling, recognition, and prediction.

Existing TSM approaches face intrinsic challenges and limitations, with one of the most critical issues being long-term dependence modelling and prediction. Very long sequences can lead to the well-known problems of gradient vanishing or exploding, significantly affecting the learning and inference of various machine learning models, including vanilla recurrent neural networks [10]. Additionally, handling

very long sequences introduces computational and memory complexity concerns for TSM [9].

To address the challenges posed by long-term dependence in time-series modeling, a GAN-based approach has been proposed in this research. Previous studies have indicated that GAN-based time-series prediction shows promise for accurate sequence forecasting [11]. It has been observed that the introduction of a discriminator in GAN highlights the capture of long-term dependence, thereby improving prediction performance at the sequence level [12]. However, the adaptation of GANs for time-series signals remains an open question, especially considering the distinct nature of time-series data compared to images. This study aims to provide a comprehensive analysis of GAN-based time series modelling, focusing on the characteristics of the generator, discriminator, model training, and performance. Additionally, a High-Dimensional Extension (HDE) block has been developed for Transformer-based models to enhance prediction accuracy and address challenges related to long-term dependencies.

B. Contribution

The contribution of this study can be summarized as follows.

- A novel RGAN Transformer architecture was developed and analyzed for accurate time-series forecasting on IVs.
- We proposed high-dimension extension block that leverage the spatiotemporal attention learning blocks to further enhance the prediction accuracy.
- We conducted extensive experiment on naturalistic data and evaluate the HDE block based on different backbone networks to show the efficiency of HD representation.

C. Paper Organization

This study is organized as follows. First, related works are discussed in Section II. In Section III, the proposed RGAN-TF model is introduced. In Section IV, model training and testing platforms are described, and extensive experiment results and comparison are proposed. Last, the paper is concluded in Section V.

II. RELATED WORKS

A. Classical TSM

Conventional statistical models such as Auto Regressive Integrated Moving Average (ARIMA), Nonlinear autoregressive exogenous (NARX) model, and exponential smoothing have shown significant efficiency in early TSM tasks, such as

*Y. Xing, XQ. Kong, and A.Tsourdos are with the School of Aerospace, Transport and Manufacturing, Cranfield University, Bedfordshire, MK43 0AL, U.K. (e-mail: yang.x, xiangqi.kong, a.tsourdos@cranfield.ac.uk).

univariate sequence modelling [13]. However, these methods have limitations in long-term sequence modelling, generation, and complex multi-variate signals, therefore, data-driven machine learning models have become the mainstream for TSM. Classical machine learning models rely on a carefully designed feature extraction on the sequence data before applying these models [14], however, such hand-crafted feature extraction rely heavily on expert knowledge in specific domains and do not help to improve the robustness and generalization of the TSM task. The end-to-end deep learning-based models, on the other hand, provides a strong improvement to conventional models. WaveNet was one branch for TSM with deep convolutional neural networks [15]. The RNN and its improved version with long-short term memory (LSTM) or gated recurrent unit (GRU) can achieve a much significant improvement and generalize very well in diverse applications such as classification, anomaly detection, forecasting, and generation [7], [16]. In [17], a probabilistic forecasting method that predicts future probability distribution was developed based on an autoregressive recurrent neural network (DeepAR).

B. Transformer-based TSM

Even though RNN-based models have achieved significant improvement for many TSM tasks, limitations in very long and high-dimension sequence modelling still exist. Recently, Transformer-based time-series models are showing the state-of-the-art results in many complex tasks such as natural language processing [18], image/video caption and alignment [19], as well as time-series modelling and forecasting on IVs [20]. It was shown that two major weaknesses (locality-agnosticity and memory bottleneck) existed in the vanilla transformer architecture when modelling the time series [21]. Therefore, multiple efficient backbone Transformer-based TSM networks were developed to overcome these drawbacks. In [22], a Reformer network was developed to further reduce the model complexity and training efficiency by introducing the locality-sensitive hashing for nearest neighbors and attention allocation in high-dimensional spaces. Based on these successful architectures, recently more advanced transformer-based models are developed. For instance, the current Informer network adopts a ProbSparse self-attention mechanism with self-attention distilling and generative decoder [9]. Autoformer, on the other hand, further improve the long-term forecasting problem with a novel decomposition architecture and Auto-Correlation mechanism [23]. Following this, currently the Frequency improved Legendre Memory model (FiLM) [24] and the temporal variation modeling-based TimesNet [25] are showing further superior and state-of-the-art (SOTA) performance in fundamental TSM tasks.

C. GAN-based TSM

Since the publication of GAN in 2014 [26], GAN has been widely applied for many computer vision tasks such as image/video generation and reconstruction. The zero-sum game between the generator and discriminator enables GAN to learn the original data distribution but also enables it

to maintain the diverse generation results of the model. Recently, GAN has been successfully adapted to the TSM task and it was found that GAN contributes to stronger long-term dependence modelling and the integration of the discriminator provides high-level guidance to the model training [27]. For example, once the GAN was published, a recurrent conditional GAN model was developed to generate near real sequence examples [28]. In [27], a Time-series GAN model was proposed to strengthen the preservation of temporal dynamics and temporal correlation in the new sequence with the original sequence by combining an unsupervised paradigm with the control afforded by supervised learning. A learned embedding space, which encourages the network to adhere to the training data is designed to jointly optimize the supervised and adversarial objectives. GAN-based TSM can be efficiently used for synthetic sequence generation and data augmentation. An implicit adversarial sequential data generation model was developed in [29]. The proposed COT-GAN model designed a Causal Optimal Transport (COT)-based loss function, which combines the classic optimal transport methods with an additional temporal causality constraint.

Considering another application of TSM, GAN-based time-series forecasting has slightly different and more complex objectives than sequence generation. Specifically, time-series forecasting requires the model to understand the temporal dynamics for both historical data and future sequences within a specific horizon. Besides, sequence forecasting can be more complex and challenging than single-point prediction due to the uncertain nature of time series. To reduce the Transformer model complexity and improve the sequence level prediction accuracy, an adversarial sparse transformer network which adopts the Sparse Transformer to learn the sparse attention map was proposed in [12].

III. METHODOLOGY

A. Overall System Architecture

In this study, we introduce a novel RGAN-based Transformer framework designed for time-series sequence forecasting. The framework comprises two fundamental modules: the generator and the discriminator within the RGANFormer architecture. The discriminator, employing a standard WaveNet, serves a dual purpose. First, it classifies between the original and generated sequences. Second, it offers high-level guidance to the generator, facilitating the learning of the genuine data distribution from the original sequence. The generator in the RGANFormer consists of four essential components: encoder, decoder, high-dimensional sequence attention, and masked reconstruction. The encoder and decoder adhere to the well-established Transformer-based architecture such as Autoformer and FiLM, which are two of the state-of-the-art (SOTA) backbones for TSM. Within this framework, the sequence convolutional block attention module (CBAM) [33] is integrated into the backbone model. This inclusion aims to learn high-dimensional features and representations from the encoder's learned feature map, extracting attentional representations crucial for supporting

predictions in the decoder. Additionally, we introduce an input reconstruction model post-encoder to reinforce the model's comprehension of the input sequence. This unique contribution module to TSM ensures that the model learns to predict future sequences accurately based on a robust understanding of the input distribution.

The overall structure of the RGANFormer framework is visually represented in Fig. 1. Formally, we define our observed variable sequence as $X = \{x_1, \dots, x_{t_0} \mid x_i \in \mathbb{R}^k\}$ and observed target sequence $Y = \{y_1, \dots, y_{t_0} \mid y_i \in \mathbb{R}^d\}$, and the future target sequence as $Y_f = \{y_{t_0+1}, \dots, y_{t_0+h} \mid y_i \in \mathbb{R}^d\}$, where both x_i and y_i can be either multivariate or univariate sequences with k and d dimensions, $t_0 \in \mathbb{N}$ is the historical observation horizon for the input sequence and $h \in \mathbb{N}$ is the prediction horizon of the model. The overall target is to learn a representation model f_m that make accurate sequence prediction of $\hat{Y}_f = \{\hat{y}_{t_0+1}, \dots, \hat{y}_{t_0+h}\}$ based on the training pair X and Y following (1) below.

$$\hat{Y}_{\rho, t_0+1:t_0+h} = f_m(Y, X; \Phi) \quad (1)$$

where $\hat{y}_{\rho, t_0+\tau} \in \hat{Y}_{\rho, t_0+1:t_0+h}$ is the ρ^{th} point-wise prediction value at time step $t + \tau$, f_m is the prediction model with learnable parameters Φ .

B. RGANFormer

The Relativistic GAN (RGAN) incorporates a "relativistic discriminator" to assess the likelihood that a given set of real data is more realistic than fake data [30]. In comparison to its non-relativistic counterparts, RGAN exhibits significantly enhanced stability, resulting in the generation of higher-quality sample data. Time series data often exhibits complex long-term and short-term temporal dependencies, with an emphasis on learning trends, cyclic patterns, and seasonal trends being crucial. Unlike standard GANs that evaluate whether an input is real or fake, RGAN is posited to be more suitable for Time Series Modeling (TSM) tasks. This is attributed to the relativistic cost function, which proves beneficial in reinforcing the modelling of long-term and short-term temporal dependencies. Furthermore, we contend that RGAN facilitates the emphasis on learning the data distribution for the generator, bringing it closer to the real data distribution. These assertions find support in the evidence presented in the comparative results in Section IV.

Specifically, we apply a relativistic average GAN (RaGAN) model to enable the discriminator to estimate the probability of the given real data is more realistic than the fake data, on average. In RaGAN, we can define the discriminator as $D(x) = \text{sigmoid}(C(x))$, where $C(x)$ is the non-transformed layer. Let The Θ^G and Θ^D be the parameters for the generator G and discriminator D. The function

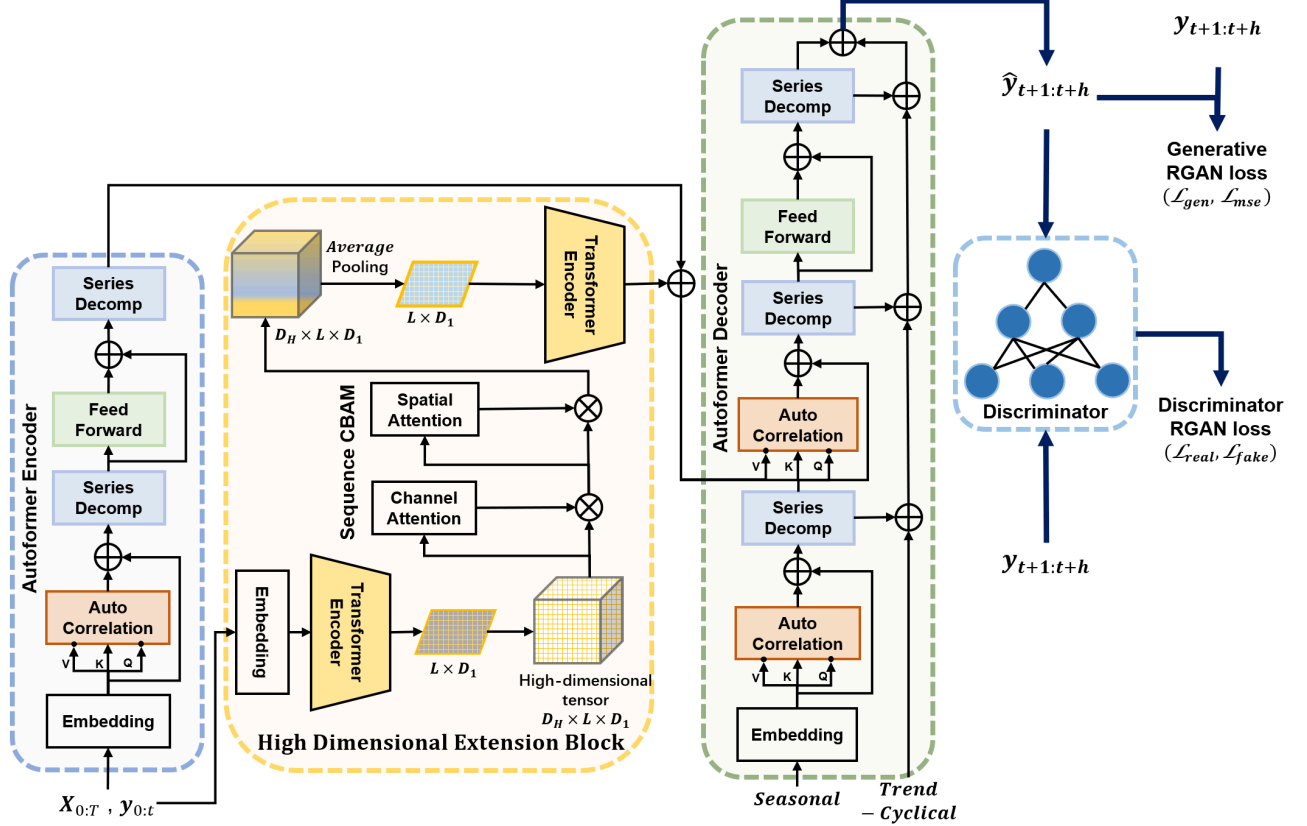


Fig. 1. The proposed RGANFormer Architecture based on Autoformer backbone network. The HDE block was integrated into the Autoformer backbone network and the relativistic GAN training framework was implemented for more accurate temporal representation.

$G(X_{1:t}, Y_{1:t}; \Theta^G)$ generate the prediction $\hat{Y}_{t+1:t+h}$, which is the input to discriminator D . Let $D(Y_{t+1:t+h}; \Theta^D)$ be the output of the discriminator, Θ^G and Θ^D are the parameters of generator and discriminator that to be learned, then the overall RaGAN loss function can be formulated as the min-max optimization problem $\arg\min_G \max_D L_{RaGAN}(\Theta^G, \Theta^D)$. We follow the same definition for the relativistic loss for RaGAN $L_{D_RaGAN}(\Theta^G, \Theta^D)$ and L_{G_RaGAN} as used in [30].

$$L_{D_RaGAN}(\Theta^G, \Theta^D) = -\mathbb{E}_{Y_r \sim \mathbb{P}}[\log(\bar{D}(Y_r))] - \mathbb{E}_{Y_f \sim \mathbb{Q}}[\log(1 - \bar{D}(Y_f))] \quad (2)$$

$$L_{G_RaGAN}(\Theta^G, \Theta^D) = -\mathbb{E}_{Y_f \sim \mathbb{Q}}[\log(\hat{Y}_f)] - \mathbb{E}_{Y_r \sim \mathbb{P}}[\log(1 - \bar{D}(Y_r))] \quad (3)$$

where $\mathbb{P}$ is the distribution of real data, $\mathbb{Q}$ is the distribution of fake data. Y_r is the real data, and Y_f is the fake data.

1) *Generator*: In this study, the generator is constructed using the Autoformer and FiLM backbone networks. The RaGAN generator loss is employed to facilitate the learning of the target data distribution. However, recognizing that this loss function alone may not efficiently address time-series data due to potential large quantile errors in the sequence, we complement it with the mean squared loss (MSE) function $MSE = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2$. This addition of the mean squared loss function aims to enhance the compensatory aspects of the adversarial loss in capturing the intricacies of time-series data.

2) *Discriminator*: Time-series signals are often susceptible to real-world noise, leading to significant uncertainty and accumulated errors. Additionally, time sequences exhibit diverse temporal dependencies and dynamics among features, including seasonal and trend-cyclical components, as explored in previous studies [31]. In this research, our objective is to leverage an adversarial training process, employing a suitable discriminator to regulate the generator outputs. This approach aims to enhance the precision of learning the data distribution at the sequence level. Addressing the adversarial training challenge between the generator and discriminator, it is crucial not to implement an overly strong discriminator. An excessively strong discriminator may hinder the efficient updating of the generator through the adversarial training process with backpropagation. Hence, we opt for the WaveNet model [15], which incorporates stacks of dilated causal convolutional layers to enable a large receptive field. This choice is made to strike a balance in the adversarial training dynamics and foster effective updates within the generator.

3) *Multi-task Loss Function Wrapper*: In this study, we use the homoscedastic uncertainty-based approach [32] to weigh multiple loss functions by considering the homoscedastic uncertainty of each task. Specifically, following the definition of the homoscedastic uncertainty weighted loss, we applied a multi-task learning approach to form the loss function L_G and L_D for the generator and discriminator respectively.

$$L_G = \frac{1}{2\sigma_1^2} L_{G_cls} + \frac{1}{2\sigma_2^2} L_{mse} + \log \sigma_1 \sigma_2 \quad (4)$$

$$L_D = \frac{1}{2\sigma_3^2} L_{D_{cls_real}} + \frac{1}{2\sigma_4^2} L_{D_{cls_fake}} + \log \sigma_3 \sigma_4 \quad (5)$$

where the cross-entropy loss function is used to calculate the typical classification loss for the real and generated samples in the generator L_{G_cls} and discriminator $L_{D_{cls_real}}$ and $L_{D_{cls_fake}}$ parts, L_{mse} is the point-wise MSE loss for the generated samples, σ_i is the learnable parameter that represents the homoscedastic uncertainty or the observation noise.

C. High-dimensional Representation

Another motivation for this study is to analyze the impact of a high-dimensional representation of the original sequence data on the long-term prediction task. The hypothesis posits that introducing a higher dimension into the original feature map of the sequential signal and learning high-dimensional attention may contribute to a better representation of long-term dependencies, thereby benefiting sequential prediction. Consequently, a sequential CBAM module is designed and seamlessly integrated into the GAN framework. This module is conceived to enhance the model's ability to capture intricate temporal dependencies for improved sequential prediction outcomes.

Specifically, given the original sequential input batch $X_i \in \mathbb{R}^{len \times dim}$, we use a fully connected layer and a Transformer encoder to learn a separate temporal representation and increase the dimension of X_i , such as $X'_i \in \mathbb{R}^{len \times 512}$.

$$X'_i = \text{Transformer_enc1}(FC(X_i)) \quad (6)$$

Next, we expand X'_i to $X''_i \in \mathbb{R}^{len \times 512 \times 1}$ and use an FC layer to map X''_i to $X_{high} \in \mathbb{R}^{len \times 512 \times 64}$, then we switch the dimensions of X_{high} to $X'_{high} \in \mathbb{R}^{64 \times len \times 512}$ and use the expanded dimension as the channel dimension for CBAM.

$$X''_{high} = M_c(X'_{high}) \otimes X'_{high} \quad (7)$$

$$f_{attn} = M_s(X''_{high}) \otimes X''_{high} \quad (8)$$

where $M_c(\cdot)$ is the channel attention module and $M_s(\cdot)$ is spatial attention module which can be represented as:

$$M_c(X'_{high}) = \sigma(MLP(AvgPool(X'_{high}))) + \sigma(MLP(MaxPool(X'_{high}))) \quad (9)$$

$$M_s(X''_{high}) = \sigma(f^{3 \times 3}([AvgPool(X''_{high}); MaxPool(X''_{high})])) \quad (10)$$

where σ denotes the sigmoid function, MLP denotes the fully-connected layers, and $f^{3 \times 3}$ represents the convolution operation with filter size of 3×3 . Then, the attentional

representation $f_{attn} \in \mathbb{R}^{len \times 512}$ after CBAM and global average along the channel dimension are passed to another Transformer encoder for feature representation.

$$f'_{attn} = \text{Transformer_enc2}(f_{attn}) \quad (11)$$

Last, we sum the feature maps f_{enc} , f'_{attn} from the backbone encoder and the HDE blocks and fed into the decoder part for final prediction.

D. Model Implementation Details

We utilized a naturalistic dataset collected from an electric vehicle testing bed. This time-series driving dataset was acquired on a chassis dynamometer featuring an electric vehicle operating under typical driving cycles. To simulate real-world urban driving scenarios, we adopted the standard New European Drive Cycle (NEDC), combining the European Union Urban Driving Cycle (ECE) and the Extra Urban Driving Cycle (EUDC). The data from the CAN bus, including vehicle speed, acceleration, braking cylinder pressures from the four wheels, motor states (speed and torque), and battery states (SOC, current, voltage), among others, was collected at a sampling frequency of 100 Hz, with around 120 minutes of driving data utilized. In total 14 signal channels are collected, including the vehicle speed, acceleration, braking cylinder pressure, and battery states, etc. A detailed description of the published data can be found in [1]. The RGANFormer is trained with ADAM [34] optimizer with an initial learning rate of $8e-5$, β_1 and β_2 equals to 0.5 and 0.99. Batch size is set to 128. All experiments are implemented in PyTorch and conducted on a single NVIDIA A6000 48GB GPU. The maximum training epoch is five due to the use of the large sequential dataset. We use the initial setup for Autoformer and FiLM backbones, including the fixed 96-input sequence length and 48-start token length. The tested prediction horizons are 48, 96, 192, and 336, respectively. For the HDE, each of the Transformer blocks contains two Transformer encoder layers and each layer has eight heads.

IV. EXPERIMENT RESULTS AND ANALYSIS

In this section, we assess the sequence forecasting performance of the proposed method across two typical tasks: multivariate and univariate forecasting, respectively. For the multivariate scenario, the input and output can be represented as $X_i \in \mathbb{R}^{96 \times 14}$ and $Y_i \in \mathbb{R}^{pred_len \times 14}$, while for the univariate case, we only use the cylinder pressure from the front left wheel for evaluation, where the input and output can be demonstrated as $X_i \in \mathbb{R}^{96 \times 1}$ and $Y_i \in \mathbb{R}^{pred_len \times 1}$. As discussed in Section III.D, there are four different prediction lengths tested in this study. The typical mean square error (MSE) and mean absolute error (MAE) are used as evaluation metrics.

A. Multivariate Time-series Forecasting

We assessed the effectiveness of the RGAN framework and the HDE block by evaluating eight different models/combinations. Specifically, RGAN_AF and RGAN_FM

indicate the integration of the Autoformer (AF) and FiLM (FM) backbone into the RGAN framework. AF only and FM only refer to using the original model and training process as described in [23] and [24], respectively.

The performance of multivariate time series forecasting is presented in Table I. In the upper section of Table I (with Autoformer backbone), the RGANFormer+HDE combination achieved the lowest MSE in the four different prediction cases. Specifically, the MSE for the 48 and 96 cases decreased by 37% and 13.4%, respectively, compared to the original AF-only model. Regarding the MAE metric, the RGANFormer+HDE model exhibited superior performance with prediction lengths of 96 and 336, while slightly higher than the RGAN-only model for the 48 and 192 cases. In the lower part of the table (with FiLM backbone), the RGANFormer+HDE also achieved the lowest MSE in all four cases. However, the MAE values were higher than those of the initial FiLM model. The introduction of the RGAN framework and HDE block shows a significant contribution to the Autoformer model, however, the RGAN only model with FiLM network does not show contributions to the 192 and 336 cases.

Fig. 2 illustrates the MSE and MAE of RGANFormer+HDE for each signal channel from the EV dataset with a 48-prediction horizon. Consistent trends are observed for both the Autoformer and FiLM models. First, focusing on the MSE, it is evident that the braking cylinder pressures represent a significant source of substantial MSE for both the Autoformer and FiLM models. And the battery states (charge and voltage) present the second major source of difficulty in achieving accurate predictions, underscoring the challenges associated with these signals. Conversely, the vehicle states

TABLE I
MULTIVARIATE LONG-TERM FORECASTING RESULTS ON THE EV DATASET WITH VARIOUS INPUT LENGTH AND PREDICTION HORIZON 48, 96, 192, 336. A LOWER MSE/MAE INDICATES BETTER PERFORMANCE. THE UPPER PART IS DERIVED WITH AUTOFORMER NETWORK, AND THE LOWER PART IS DERIVED WITH THE FiLM NETWORK.

Models	Metrics	Prediction Length			
		48	96	192	336
RGAN_AF + HDE	MSE	0.0947	0.1712	0.2575	0.3175
	MAE	0.1138	0.1127	0.1868	0.1976
RGAN_AF	MSE	0.1372	0.1890	0.2620	0.3245
	MAE	0.1046	0.1299	0.1587	0.1859
AF + HDE	MSE	0.1463	0.1983	0.2621	0.3240
	MAE	0.1091	0.1283	0.1635	0.2046
AF only	MSE	0.1519	0.1978	0.2627	0.3199
	MAE	0.1136	0.1232	0.1624	0.2136
RGAN_FM + HDE	MSE	0.0561	0.1199	0.1953	0.2605
	MAE	0.0586	0.1023	0.1574	0.1956
RGAN_FM	MSE	0.0572	0.1202	0.1969	0.2657
	MAE	0.0524	0.1006	0.1587	0.1967
FM + HDE	MSE	0.0576	0.0119	0.1966	0.2595
	MAE	0.0510	0.0810	0.1191	0.1571
FM only	MSE	0.0581	0.1206	0.1960	0.2621
	MAE	0.0516	0.0815	0.1212	0.1591

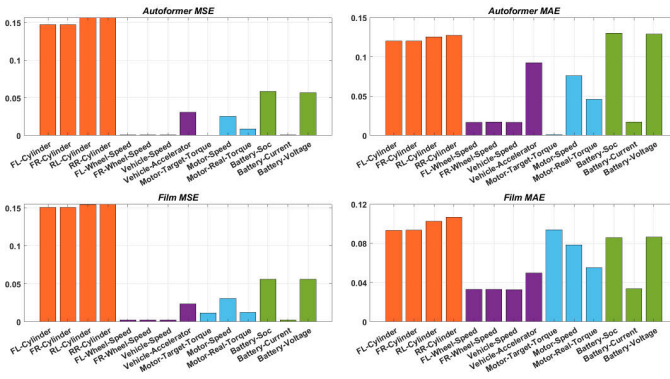


Fig. 2. Illustration of MSE and MAE for multivariate signal forecasting with Autoformer (Upper) and FiLM (bottom) backbone. The signals come from four main resources, namely, braking cylinder (orange), vehicle states (purple), motor states (blue), and battery states (green). FL, FR, RL, RR represent front-left, front-right, rear-left, and rear-right, respectively.

and motor states exhibit lower MSE, indicating a relatively easier predictability for these signals. Second, in terms of the MAE, the cylinder pressure and battery states continue to be the two predominant sources contributing to substantial prediction errors. Notably, the vehicle states demonstrate observable and competitive MAE with the Autoformer backbone. In the FiLM case, the motor states become the second significant source of elevated MAE. These consistent trends across different signal channels underline the complexities involved in accurately predicting certain signals, providing valuable insights for future model improvements.

B. Univariate Time-series Forecasting

Univariate forecasting is assessed based on the sequential data from the braking cylinder of the front left wheel. In the upper part of Table 2, with the Autoformer backbone, the RGANFormer+HDE achieved a significant decrease in MSE in all four cases compared to the original AF-only model. For the 48-prediction length, the RGANFormer+HDE exhibited a 32.2% reduction in MSE and an 18.6% reduction in MAE. The RGANFormer+HDE achieved a lower MAE in the 48 and 336 cases. Compared to the original AF, the introduction of the RGAN framework and HDE block also showed improvement in the MSE case, indicating fewer large outliers.

The bottom part of the table presents the results with the FiLM backbone. Although, as in the multivariate case, the performance improvement is less pronounced compared to the Autoformer case, and the MAEs are higher than the original model, the MSEs show an overall reduction in the four scenarios. Specifically, compared to the original FiLM model, RGANFormer+HDE reduces the MSE by 8.7%, 12.3%, 10.9%, and 7.6% in the four different cases, respectively.

Based on Tables 1 and 2, it can be concluded that:

- **Universal Improvement in MSE with RGANFormer+HDE Models:** The RGANFormer+HDE models, whether using Autoformer or FiLM backbone, show a consistent and universal improvement in MSE

TABLE II

UNIVARIATE LONG-TERM FORECASTING RESULTS ON THE EV DATASET WITH VARIOUS INPUT LENGTH AND PREDICTION HORIZON 48, 96, 192, 336. A LOWER MSE/MAE INDICATES BETTER PERFORMANCE. THE UPPER PART IS DERIVED WITH AUTOFORMER NETWORK, AND THE LOWER PART IS DERIVED WITH THE FiLM NETWORK

Models	Metrics	Prediction Length			
		48	96	192	336
RGAN AF + HDE	MSE	0.2695	0.4806	0.5609	0.7065
	MAE	0.1626	0.2335	0.2995	0.2795
RGAN AF	MSE	0.3995	0.4822	0.6200	0.7218
	MAE	0.1882	0.2376	0.2769	0.2603
AF + HDE	MSE	0.3915	0.5016	0.6260	0.7041
	MAE	0.1678	0.2133	0.2654	0.3320
AF only	MSE	0.3980	0.5051	0.6268	0.7230
	MAE	0.1929	0.1903	0.2252	0.3002
RGAN_FM + HDE	MSE	0.1405	0.2714	0.4112	0.5132
	MAE	0.1017	0.1863	0.2505	0.3225
RGAN_FM	MSE	0.1449	0.2844	0.4383	0.5286
	MAE	0.1010	0.1883	0.2783	0.3119
FM + HDE	MSE	0.1536	0.3078	0.4565	0.5523
	MAE	0.0842	0.1318	0.1818	0.2174
FM only	MSE	0.1539	0.3096	0.4617	0.5558
	MAE	0.0815	0.1329	0.1788	0.2176

across different prediction lengths. This suggests that the integration of RGANFormer+HDE models with either Autoformer or FiLM backbone is effective in enhancing performance.

- **Limited Contribution to MAE Reduction:** Though the HDE block can slightly reduce the MAE, the integration of RGAN and HDE does not lead to a significant reduction in MAE when compared to the original Autoformer and FiLM models. This is especially notable in the case of the FiLM model. Therefore, there is a suggestion for future work to conduct more channel-wise analysis and explore additional backbone models to further evaluate and improve MAE.
- **Efficiency of RGAN and HDE Blocks for MSE Improvement:** It is concluded that either the RGAN or HDE block can be effectively utilized to enhance model performance, specifically in terms of MSE. This efficiency holds true across different models and tasks, indicating the robustness and versatility of the proposed methods for various scenarios.

In summary, the study suggests the effectiveness of RGANFormer+HDE models with Autoformer or FiLM backbone in improving MSE universally, while also highlighting the need for further exploration and analysis, especially in terms of MAE reduction and the potential use of more backbone models for evaluation

V. CONCLUSION

In this study, we developed a RGAN-based adversarial training framework for time-series signal forecasting in intelligent vehicles. We believe that accurate predictions for key vehicle states can enhance various onboard functions, such

as active safety, energy management, and cybersecurity. Introducing a new high-dimensional extension block for more accurate temporal pattern representation, we combined the RGAN and HDE block to demonstrate increased prediction accuracy for both multivariate and univariate scenarios. Our results show that both the proposed RGAN framework and the HDE block are independent of specific models, and we tested their efficiency with two different state-of-the-art baselines (Autoformer and FiLM). The findings indicate that the proposed method can adapt to different models and contribute to more accurate long-term time-series forecasting. Future work will focus on 1) developing more energy-efficient independent blocks for time-series modelling, 2) integrating explainable blocks for more complex signals, and 3) conducting system testing using more challenging real-world scenarios.

REFERENCES

- [1] Y. Xing, Z. Hu, P. Hang, and C. Lv, "Learning from the dark side: A parallel time series modelling framework for forecasting and fault detection on intelligent vehicles," *IEEE Transactions on Intelligent Vehicles*, 2023.
- [2] H. Qin, M. Yan, and H. Ji, "Application of controller area network (can) bus anomaly detection based on time series prediction," *Vehicular Communications*, vol. 27, p. 100291, 2021.
- [3] Y. Xing, C. Lv, D. Cao, and C. Lu, "Energy oriented driving behavior analysis and personalized prediction of vehicle states with joint time series modeling," *Applied Energy*, vol. 261, p. 114471, 2020.
- [4] S. Zhou, J. Gong, G. Xiong, H. Chen, and K. Iagnemma, "Road detection using support vector machine based on online learning and evaluation," in *2010 IEEE intelligent vehicles symposium*. IEEE, 2010, pp. 256–261.
- [5] M. M. Ahmed and M. A. Abdel-Aty, "The viability of using automatic vehicle identification data for real-time crash prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 459–468, 2011.
- [6] O. L. Junior and U. Nunes, "Improving the generalization properties of neural networks: an application to vehicle detection," in *2008 11th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2008, pp. 310–315.
- [7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [9] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.
- [10] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*. Pmlr, 2013, pp. 1310–1318.
- [11] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A systematic literature review," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1–31, 2023.
- [12] S. Wu, X. Xiao, Q. Ding, P. Zhao, Y. Wei, and J. Huang, "Adversarial sparse transformer for time series forecasting," *Advances in neural information processing systems*, vol. 33, pp. 17 105–17 115, 2020.
- [13] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International journal of forecasting*, vol. 22, no. 3, pp. 443–473, 2006.
- [14] A. Jahangiri and H. A. Rakha, "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," *IEEE transactions on intelligent transportation systems*, vol. 16, no. 5, pp. 2406–2417, 2015.
- [15] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.
- [16] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [17] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "Deepar: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.
- [18] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023.
- [19] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.
- [20] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," *arXiv preprint arXiv:2301.07945*, 2023.
- [21] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," *Advances in neural information processing systems*, vol. 32, 2019.
- [22] N. Kitaev, Ł. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," *arXiv preprint arXiv:2001.04451*, 2020.
- [23] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 419–22 430, 2021.
- [24] T. Zhou, Z. Ma, Q. Wen, L. Sun, T. Yao, W. Yin, R. Jin *et al.*, "Film: Frequency improved legendre memory model for long-term time series forecasting," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 677–12 690, 2022.
- [25] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186*, 2022.
- [26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [27] J. Yoon, D. Jarrett, and M. Van der Schaar, "Time-series generative adversarial networks," *Advances in neural information processing systems*, vol. 32, 2019.
- [28] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," *arXiv preprint arXiv:1706.02633*, 2017.
- [29] T. Xu, L. K. Wenliang, M. Munn, and B. Acciaio, "Cot-gan: Generating sequential data via causal optimal transport," *Advances in neural information processing systems*, vol. 33, pp. 8798–8809, 2020.
- [30] A. Jolicœur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.
- [31] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.
- [32] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.
- [33] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.



DSPACE

<https://dspace.org/>

RGANFormer: Relativistic Generative Adversarial Transformer for time-series signal forecasting on intelligent vehicles

Xing, Yang; Kong, Xiangqi; Tsourdos, Antonios

2024-07-15

Institute of Electrical and Electronics Engineers (IEEE)

<https://doi.org/10.1109/IV55156.2024.10588746>

Xing Y, Kong X, Tsourdos A. (2024) RGANFormer: Relativistic Generative Adversarial Transformer for time-series signal forecasting on intelligent vehicles. In: 2024 IEEE Intelligent Vehicles Symposium (IV), 2-5 June 2024, Jeju Island, Republic of Korea, pp. 3241-3247