IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 April 2023, accepted 28 June 2023, date of publication 6 July 2023, date of current version 31 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3292911

RESEARCH ARTICLE

Non-Intrusive Load Monitoring Based on Residual U-Net and Conditional Generation Adversarial Networks

JINLONG WANG[1], (Graduate Student Member, IEEE), CHENGXIN PANG[1], (Member, IEEE), XINHUA ZENG[2], (Member, IEEE), AND YONGBO CHEN[3]

[1]School of Electronics and Information Engineering, Shanghai University of Electric Power, Shanghai 200000, China
[2]Academy for Engineering and Technology, Fudan University, Shanghai 200433, China
[3]ZTE Corporation, Shenzhen 518057, China

Corresponding author: Chengxin Pang (chengxin.pang@shiep.edu.cn)

This work was supported by the State Grid Corporation of China Science and Technology Project "Research on Service Driven Planning Method and Key Technologies of Electric Sensor Network" under Grant SGSCJY00GHJS2000014.

ABSTRACT A non-intrusive load disaggregation method based on residual U-Net and conditional generation adversarial networks (RUCGAN) model is proposed to address the low decomposition accuracy and poor generalization of traditional load disaggregation algorithms. The method is based on a conditional generative adversarial networks (CGAN), which is a variant of the encoder-decoder model that is suitable for processing time-series data and overcomes the limitation of requiring a manually designed feature extractor in traditional encoder-decoder structures. By introducing the U-Net structure as the encoder of the CGAN network, the size of the feature map can be gradually reduced through convolution and pooling operations, and gradually restored through deconvolution and upsampling operations. The U-Net structure also has skip connections that effectively preserve feature information and accelerate gradient propagation, thus improving model stability and generalization. Furthermore, combining the residual structure with the U-Net structure further enhances the model's performance, as the residual connections can effectively reduce the number of network parameters and computation. Experimental results show that the MAE value of the model on the UK-DALE dataset decreased by at least 20.5%, and the MAE value of the model on the REFIT dataset decreased by at least 9.9%. Moreover, while improving the decomposition accuracy, the model size decreased by at least 5.6%.

INDEX TERMS Conditional generative adversarial networks, NILM, ResU-net.

I. INTRODUCTION

With the rapid development of society, energy sources such as oil are being used in large quantities, thus causing energy stress. The burning of fossil fuels is contributing to global warming. Energy is a significant constraint on human development, whether in the past or the future. The use of renewable energy sources, such as wind, water, and solar energy, optimizing the structure of energy use and increasing energy use efficiency are two of the main sustainable ways to save energy. With residential and commercial buildings consuming 36% of their energy and accounting for 39% of the world's carbon emissions, energy use must be managed efficiently to improve energy efficiency. According to research reports, energy losses can be reduced by up to 20% with the feedback of load consumption data to the user [1]. In recent years, smart grid technologies based on artificial intelligence have received close attention from researchers at home and abroad. Load Monitoring (LM) technology is an important part of the smart grid and is an effective way to reduce losses and energy consumption in the grid. There are two types of load monitoring: hardware-centric Intrusive Load Monitoring (ILM) and software-centric Non-intrusive Load Monitoring (NILM). While traditional ILM obtains more comprehensive

The associate editor coordinating the review of this manuscript and approving it for publication was Moussa Ayyash.

and accurate measurement data, it does so at the price of high cost and low privacy [2]. In contrast, NILM can measure and analyze the power consumption of load devices by simply installing smart metering devices at the entrance to the house power bus. This guides customers to use electricity rationally and more efficiently and helps the power company analyze customers' electricity consumption patterns. Non-intrusive load monitoring was first proposed by Professor Hart in 1991 to identify load appliances using a supervised classification model combined with a clustering approach on the P-Q plane [3]. In terms of load disaggregation, Professor Hart proposed the Combinatorial Optimization (CO) algorithm, which has poor disaggregation performance because the CO algorithm cannot distinguish between appliances with similar load characteristics. In 2011, [4] proposed the Hidden Markov Model (HMM) algorithm, which transforms the load disaggregation problem into a statistical and probabilistic modelling problem to identify individual loads from the total power consumption. In the same year, [5] proposed four variants of HMM, Factorial HMM (FHMM), Conditional FHMM (CFHMM), Factorial Hidden semi-Markov Model (FHSMM) and Conditional Factorial Hidden semi-Markov Model (CFHSMM). Modeling and experimenting with these four models, CFHSMM outperforms the remaining three models, but the complexity is too high. In 2013, [6] proposed a load disaggregation method based on Bayesian networks and Hidden Markov models with good results. In the same year, [7] proposed an FHMM-based algorithm that achieves better results for multi-state loads by selecting multiple load features in series for modeling and training. Parson et al. [8] cited the difference HMM proposed in [9] to construct a generic model based on Viterbi and expectation-maximization algorithms. The model only performs well on periodic devices; thus, the model is not very scalable. Algorithms based on HMM and various variants such as the Adaboost algorithm [10], K-nearest neighbor algorithm [11], support vector machine algorithm [12] and fuzzy algorithm [13], [14], [15] have a wide range of applications in disaggregation tasks. Still, they are more dependent on a priori knowledge modeling and are generally effective for multi-state loads.

Recently, deep learning algorithms have been widely adopted in fields such as computer vision [16], speech recognition [17], and natural language processing [18], because they can extract the inherent features of raw data. This has led many researchers to conduct related studies on energy disaggregation based on deep learning. Kelly et al. [19] first proposed the application of deep learning methods to load disaggregation, and tested recurrent neural networks (RNNs), denoising autoencoders (DAEs), and a regression model that outputs device on/off times and average power on the UK-DALE public dataset. They obtained performance that was comprehensively superior to CO and FHMM algorithms. Reference [20] proposed a deep neural network model based on DAE and improved Kelly's data processing method, which was validated on multiple datasets and achieved better performance than Reference [19]. Reference [21] proposed an energy disaggregation method based on the variational autoencoder framework, which accurately generated more complex load distributions, thereby improving the reconstruction of power signals for multi-state appliances. Reference [22] designed an encoder-decoder model with an attention mechanism, which is a commonly used method in sequence-to-sequence models to focus on information related to output in longer input sequences. The performance on the REDD and UK-DALE datasets indicated that the model has stronger generalization ability and higher accuracy. Although models based on the encoder-decoder architecture have shown superior performance in handling time series problems compared to other algorithms, such structures typically require a manually designed feature extractor to extract features, and cannot utilize prior knowledge or knowledge from other domains to assist in generating load signals. This can result in generated signals that are not realistic enough compared to actual load signals.

The sequence-to-point model proposed in [23] predicts only the midpoint of the input sequence window, resulting in smaller errors and lower computational complexity compared to the sequence-to-sequence method that predicts the entire output sequence. Reference [24] proposes a sequence-to-point model based on time convolutional neural networks that provide a more flexible receptive field. Seq2subseq, introduced in [25], is a hybrid of seq2point and seq2seq that reduces over-prediction of the output sequence while maintaining contextual relevance. Reference [26] proposes a novel LSTM-based algorithm that integrates seq2point and LSTM to reduce computation while ensuring effective feature extraction. Seq2point predicts the electrical state of each time step by treating the entire time series as input, while seq2seq predicts the electrical state of the entire time series. In contrast, seq2subseq is more suitable for NILM because it decomposes the time series into sub-sequences and independently performs classification in each sub-sequence, which better handles changes in electrical state. Additionally, seq2subseq can share features between sub-sequences, enhancing the accuracy and generalizability of the model.

In this paper, we apply the CGAN [27] model to non-intrusive load disaggregation and introduce deep residual U-net [28]. The ResU-Net was initially applied in the field of image segmentation in computer vision. Image segmentation is the task of assigning each pixel in an image to one of several pre-defined categories. Similarly, NILM aims to identify and predict the electrical behavior of individual appliances through monitoring the entire building or home energy information. Both require analysis and prediction of complex multidimensional data, thus they share some similarities. Representing energy data as a 2D matrix and processing it using image segmentation-based methods holds some research value. The residual network can deepen the network depth without causing gradient disappearance or gradient explosion problems. The main contributions of this paper are as following:

- We propose a new non-intrusive load disaggregation algorithm called RUCGAN. Compared to traditional encoder-decoder structures, RUCGAN can handle data with conditional information, help the generator generate more realistic load signals, and automatically learn feature representations.
- To improve the decomposition accuracy of the model, we used a U-Net network as the encoder. This structure can effectively extract signal features, and uses skip connections to transmit high-level semantic information to the decoder, thereby helping the decoder to better recover the output.
- Introducing residual blocks and utilizing residual connections can effectively reduce the number of network parameters and computational complexity, while avoiding the problem of model degradation and making the model more stable.
- The experimental results have proven the effectiveness of the proposed algorithm. On the UK-DALLE and REFIT datasets, the MAE value of the algorithm decreased by at least 20.5% and 9.9%, respectively. Moreover, from the decomposition results, the algorithm has a good fitting degree, which verifies its generality and effectiveness.

II. METHODOLOGY FOR NILM

The concept of NILM was introduced by Hart, and the systematic NILM process he proposed has been used to this day. Since then, the NILM research has attracted the attention of related researchers and spawned numerous research results. The disaggregation of the load appliance can be abstracted as a mathematical problem. This disaggregation problem can be formulated to show that: assuming that there are N electrical devices in residence, their sampling time series are $t = [1, 2, \cdots, T]$, and the total power consumption series are $p = [p_1, p_2, \cdots, p_T]$, the total power at the moment of entrance t can be expressed as follows:

$$p_t = \sum_{i=1}^{N} (a_i p_t^i) + \sigma(t) \tag{1}$$

where a_i represents the switch operation status of the ith device. If the operation status is on, then $a_i = 1$, otherwise $a_i = 0$. p_t^i is the operating power of the ith device, $i \in [1, 2, \cdots, N]$. $\sigma(t)$ represents the noise generated by the power fluctuations of the equipment and the environment.

The task of the NILM is to infer the power contribution of the device at moment t. The problem of solving NILM can be transformed into the problem of solving the value of a_i for different devices at different moments. Specifically, $\sigma(t)$ represents the noise generated by the environment, which should have a small value in theory. It is assumed that in the case of incorrect determination of the equipment operating state, the first part of p_t, or $\sum_{i=1}^{N}(a_i p_t^i)$. Its difference from the actual prediction is attributed to $\sigma(t)$, which leads to a larger $\sigma(t)$. Therefore, (1) is modified to transform the problem into one of solving for the optimal equipment operating state $a = [a_1, a_2, \cdots, a_N]$, which is expressed as follows:

$$\hat{a} = \arg\min_a \sigma(t) = \arg\min_a p_t - \sum_{i=1}^{N} (a_i p_t^i) \tag{2}$$

Obviously, the above problem is an NP-hard problem. This means that it is computationally intractable and difficult to solve precisely by impractical exhaustive techniques. As computer technology has evolved, many new approaches to address such problems have emerged, such as solving the problem through machine learning.

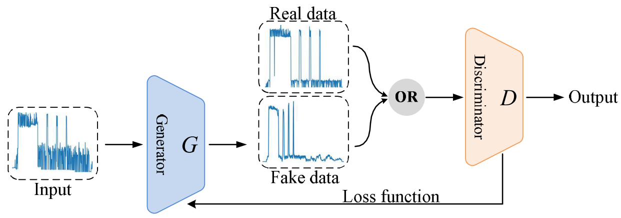

FIGURE 1. Overall structure of RUCGAN.

III. PROPOSED METHOD

The overall structure of this study is shown in Fig. 1. The following will describe the proposed RUCGAN, including the U-Net module, the residual module, and the loss function. The U-Net module and the residual block together constitute the generator module.

A. LOAD DISAGGREGATION PRINCIPLE OF THE RUCGAN MODEL

GAN [29] has excellent generation ability and data augmentation ability, and it has been widely used in fields such as image generation, natural language processing, and speech synthesis. The main idea of using GAN for NILM problem is to use generative adversarial learning to decompose the entire electricity consumption data into the usage data of individual appliances, thus achieving appliance detection and energy consumption monitoring. Through this approach, GAN can effectively solve the NILM problem. We optimized the original GAN framework as follows: using the power signal of the load appliance as conditional information to guide the output of the generator, using the U-Net which is good at handling time series as the generator of the model, and fusing residual blocks with U-Net to improve the transfer of feature information..

RUCGAN takes the aggregated power signal and appliance power signal as inputs. By utilizing CGAN, it can automatically learn the relevant feature representations from the conditional information and generate more accurate appliance disaggregation signals. The output of the generator is mixed with the real appliance power signal and used as input for the discriminator. The discriminator can then judge the

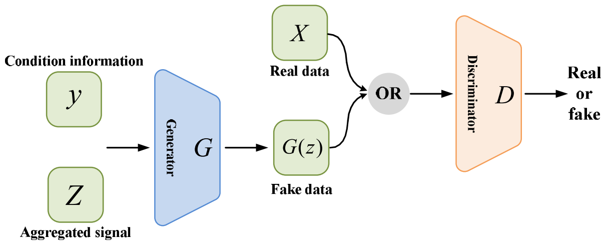

FIGURE 2. CGAN model framework.

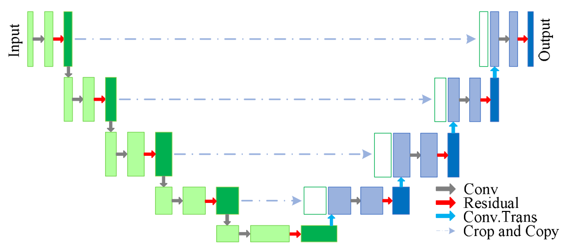

FIGURE 3. Generator module structure.

difference between the generated and real power signals, which guides the updating of the generator. Through repeated iterations, the generator can generate more realistic individual appliance usage data, thereby achieving disaggregation of the aggregated power signal.

B. ADD CONDITION INFORMATION

During the training process, GAN lacks a clear objective function and suffers from mode collapse and mode dropping problems, which can lead to generated load signals that are not realistic or lack diversity. Therefore, conditional information is crucial for load disaggregation tasks. To address these issues, we use the power signal of the target appliance as conditional information to guide the generator's data generation, as shown in Fig. 2.

In this approach, the target appliance's power signal y is paired with the aggregate power signal Z to produce a conditioned load signal, which is then input to the generator G. G generates a fake target appliance power signal $G(z)$, which is input along with the true target appliance power signal X to the discriminator D. D computes the error between $G(z)$ and X and feeds the data back to G to guide its updates. Through iterative training, the model can gradually learn the features of the appliance power signal and effectively predict and separate energy consumption.

The optimization objective function of GAN is shown in (3):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_Z(z)}[\log(1 - D(G(z)))] \quad (3)$$

where $D(x)$ denotes the probability that the true sample x is discriminated as a true sample by the discriminator, $D(G(z))$ denotes the probability that the generated sample data is discriminated as a true sample by the discriminator. After introducing conditional information y, the objective function is as shown in (4):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x\,|y)] + E_{z \sim P_Z(z)}[\log(1 - D(G(z\,|y)))] \quad (4)$$

From equation (4), it can be seen that the objective function of CGAN is equivalent to a "min-max optimization" problem, which can be achieved in two steps. The first step optimizes the discriminator model D, and the second step optimizes the generator model G, which is expressed as equations (5) and (6).

$$\max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x|y)] + E_{z \sim P_Z(z)}[\log(1 - D(G(z|y)))] \quad (5)$$

$$\min_G V(D, G) = E_{z \sim P_Z(z)}[\log(1 - D(G(z|y)))] \quad (6)$$

We define the label for real samples as 1 and the label for generated samples as 0. The goal of D is to distinguish between real and fake samples. We expect the value of $D(x|y)$ to be as close to 1 as possible and the value of $D(G(z|y))$ to be as close to 0 as possible. Based on these two conditions, we optimize D using sample data to maximize $V(D, G)$, which is the meaning of formula (5). Similarly, formula (6) shows that we need to optimize $D(G(z|y))$ to be as close to 1 as possible, which is equivalent to minimizing $V(D, G)$. By iteratively training the generator and discriminator, we can obtain the global optimal solution when $p_{data} = p_z$.

C. ENCODER MODULE

In the NILM problem, the encoder of the CGAN network usually adopts models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). Common CNN models include VGG and ResNet, which are designed for image classification and object detection tasks. Although they perform well in the image domain, these models usually ignore temporal information and cannot handle sequential data. Common RNN models include LSTM and GRU, which are often used for modeling sequential data. However, they may be limited in modeling long-term dependencies and may face issues such as vanishing or exploding gradients, making them difficult to train.

To address the aforementioned issues, we designed the encoder using U-Net [30]. U-Net has the advantage of extracting multiscale, high-level semantic information and has strong feature reuse capabilities, which is particularly useful for load disaggregation. Moreover, U-Net has a similar encoder-decoder structure, which can preserve the details of the input signal, enabling it to achieve better performance in load disaggregation.

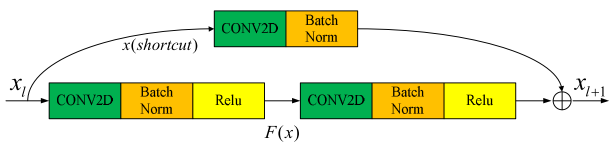

FIGURE 4. Residual module diagram.

The encoder framework is shown in Fig. 3. The left half of the figure is the encoding layer, which is used to gradually extract features from the time series and reduce the resolution of the feature map. The right half is the decoding layer, which is used to restore the feature map and generate the output result. In addition, there are some skip connections between the encoder and decoder to fuse the low-level and high-level feature maps, thereby retaining more original information and improving the accuracy of the model for load disaggregation.

D. FUSION OF RESIDUAL BLOCKS

In U-Net, multiple downsampling and upsampling operations may cause gradient vanishing. This is because during the backpropagation process, gradients need to be propagated from high-level layers to low-level layers. Each downsampling operation reduces the size of the feature map by half, which means that gradients need to pass through multiple levels of downsampling and upsampling operations to reach shallow layers. To address this issue, we use residual blocks to directly add the information of the original input to the output. This method can help the network retain more original information during the sampling process and improve the stability of the model.

The residual block is a network module proposed by He et al. [31] in ResNet, which includes a shortcut connection and a convolution block. The shortcut connection directly adds the input to the output, thus preserving the original information. The convolution block is responsible for feature extraction. The structure of the residual block is shown in Fig. 4.

The main path is $F(x)$, the shortcut is x, and the output is:

$$x_{l+1} = x_l + F(x_l, W_l) \tag{7}$$

By recursion, the feature expression of any deep residual module can be obtained, which is expressed in follows:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \tag{8}$$

For backpropagation, assuming ε as the loss function, (9) can be derived according to the chain rule of backpropagation.

$$\frac{\partial \varepsilon}{\partial X_l} = \frac{\partial \varepsilon}{\partial X_L}\frac{\partial X_L}{\partial X_l} = \frac{\partial \varepsilon}{\partial X_L}(1 + \frac{\partial}{\partial X_L}\sum_{i=l}^{L-1} F(X_i, \omega_i)) \tag{9}$$

From (9) we can find that no matter how small the derivative parameter in the brackets is, the presence of the constant 1 and the fact that the concatenated multiplication in the original chain derivative becomes concatenated addition guarantees that the node will not experience gradient disappearance or gradient explosion.

The framework of the fusion between residual block and U-Net is shown in Fig.4, where the dark module represents the residual structure.

E. LOSS FUNCTION

From (4), the objective function of CGAN is

$$\begin{aligned} &L_{CGAN}(G, D) \\ &\quad = E_{x\sim P_{data}(x)}[\log D(x\,|y)] + E_{z\sim P_Z(z)}[\log(1 - D(G(z\,|y)))] \end{aligned} \tag{10}$$

The L1 loss function is added to the objective function to improve the generation quality of the generator. The L1 loss function is expressed as

$$L_{L1}(G) = E_{x\sim p_{datd}(x), z\sim p_Z(z)}[\|y - G(z\,|y)\|_1] \tag{11}$$

The traditional loss function of CGAN uses L2 distance instead of L1, but the L1 loss function has strong robustness. The L1 loss function is not susceptible in the face of observations with significant errors. This is because the L1 loss function adds only an absolute value error, while the L2 loss adds the square of the error. When the error is large, we need to adjust the model more to fit this observation, so the L2 loss function is less stable than the L1 loss function.

The final objective function is obtained by combining (10) and (11). The formula is as follows:

$$G^* = \arg\min_G \max_D L_{CGAN}(G, D) + \lambda L_{L1}(G) \tag{12}$$

where λ is the coefficient of L1 loss function.

IV. EXPERIMENTS

We compare the model with LSTM [19], VAE [21], Seq2point [23], TCN [24], seq2subseq [25], LSTM+ [26] and ResNet+ [32], to verify the validity and superiority of this model in terms of power disaggregation accuracy. To ensure fairness, we make the experimental process as consistent as possible. The proposed model is implemented in Python using TensorFlow and is trained on NVIDIA RTX 2080 super.

A. DATA SET

In this paper, we choose public datasets UK-DALE [33], and REFIT [34]. Choosing two datasets from different regions can be a good way to verify the generalization of the proposed model. We select kettle, fridge, washing machine, microwave and dishwasher as the target appliances for disaggregation. The data used for training and testing do not come from the same residence and can satisfy the requirement of deep learning generalizability. In addition, these five target appliances have different operating states. The two loads, kettle, and refrigerator, have only two operating states, and the refrigerator is cyclical. Load changes in washing machines and dishwashers are characterized by multiple states and long periods. The microwave oven has a short

TABLE 1. Building used for training and testing(UK-DALE).

Appliance	Training house	Testing house
Kettle	1,3,5	2
Fridge	1,4,5	2
Washing machine	1,5	2
Microwave	1,5	2
Dishwasher	1,5	2

TABLE 2. Building used for training and testing(REFIT).

Appliance	Training house	Testing house
Kettle	3,4,5,6,7,8,9,11,17,19,20	2
Fridge	2,5,9,12,19	15
Washing machine	1,2,3,5,7,9,10,15,16,17,18,19,20	8
Microwave	2,3,6,10,11,12,15,17,19,20	4
Dishwasher	1,2,3,5,7,9,10,16,18	20

running time and frequent power changes. Thus, the disaggregation performance of the model proposed in this paper can be comprehensively verified.

UK-DALE records load data, including active power, current information and appliance switching status, for five UK households through devices such as current sensors and power loggers. The sampling period lasted from November 2012 to January 2015, and the sampling frequency was 1/6 Hz. The load data for each residence always contains one master meter power record and a varying number of individual power records for each appliance.

REFIT dataset is currently the largest dataset in the NILM research area, and it contains data from 20 houses in the UK between 2013 and 2015, with data sampled at a frequency of 8 seconds. The dataset provides us with a variety of different electricity usage behaviors, enriching the diversity of the data.

B. DATA PROCESSING

To verify the generalization performance of the model, we set up the training and test sets for the five target appliances, as shown in Table 1 and Table 2. The load data activation extraction of the target appliances is performed with the help of the NILMTK [35] toolkit proposed by Batra et al. The power thresholds of the five appliances are obtained as shown in Table 3. The power threshold value allows us to know the switching state of the target appliance and prepares the ground for the calculation of the evaluation index below.

C. METRICS

NILM's programs require appropriate performance evaluation metrics to measure the program's strengths and weaknesses. Hart used the percentage of correctly classified power events to evaluate his program in his pioneering work. This section describes the evaluation metrics of common disaggregation algorithms and points out their characteristics. Event detection is essentially a dichotomous problem that aims to determine the presence or absence of events. But the load disaggregation based on event classification is a multi-classification problem. Therefore the relevant statistical evaluation metrics in classification problems are often used to assess the performance of the NILM algorithm [36]. In statistics, four indicators are used to describe the relationship between predicted outcomes and actual values in dichotomous problems: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

TABLE 3. Power threshold of the target appliances.

Appliance	On power threshold (Watts)
Kettle	2000
Fridge	50
Washing machine	20
Microwave	200
Dishwasher	10

A series of relevant evaluation metrics can be obtained from four base metrics, mainly including precision, recall, accuracy, and F1-Score.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{16}$$

The above four evaluation metrics evaluate the model mainly in terms of accuracy, and in addition, we also introduce the MAE metric to evaluate our model in terms of total power. MAE is calculated as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\hat{y}_t - y_t| \tag{17}$$

where $\hat{y}_t$ denotes the disaggregation value of the model at moment t and y_t denotes the true value at moment t, and T denotes the total amount of predicted time points. MAE characterizes the average error between the target electrical power decomposed by the model and the actual electrical power consumed.

V. RESULTS

Based on the experiments, we will analyze the results in three aspects: training model size, visual analysis of load disaggregation results, and evaluation metrics.

A. MODEL SIZE

The Seq2subseq model uses the U-Net structure to build a code-and-decode framework and the number of Encoder-Decoder layers of the model is 8. While the Encoder-Decoder layer of our proposed model is 5, the model depth is reduced but the disaggregation performance of the

TABLE 4. Model parameters.

Method	Total parameters	Model size (MB)
Seq2point	30708249	351
TCN	18580529	212
Seq2subseq	14315778	158
RUCAN	**13514004**	**149**

model is improved. This is attributed to the ResU-Net, where we add the residual module to each Encoder-Decoder layer, which can be seen in Fig. 3. From Table 4, it can be seen that the number of parameters in the RUCGAN model is nearly 800,000 less than seq2subseq, and the model size is reduced by 5.6%. The model size was reduced by 57.5% and 29.7% compared to Seq2point and TCN, respectively.

B. VISUAL ANALYSIS OF LOAD DISAGGREGATION RESULTS

From Fig. 5, we can conclude that the operating states of the fridge, dishwasher and washing machine are more complex, and the running time of the washing machine is relatively long. The microwave oven and kettle are in a simple state of operation and have a short working time. But these two appliances operate more frequently. For the fridge, it can be seen from the diagram that it has two operating states, one is a long sequence shaped like a square, and the other is a short sequence shaped like a rectangle. The power fit of the short sequences is higher than that of the long sequences, mainly because a wave with a short duration and large amplitude appears in each long sequence. The model does not fit the spikes well, but the RUCGAN model gives a high-accuracy disaggregation for both start-end times and peaks. For the washing machine, a complete section of the waveform is given in Fig. 5, and the fit of the disaggregation results remains good despite the long time span. Whether it is the main wave on the far left or the three secondary waves on the right, and even if the wave is very dense on the far right, the model can still decompose its contours. For microwave ovens and kettles, which are two appliances with simple operating states, the accuracy of the model fit is high, especially for kettles, where the model provides a perfect load disaggregation even when the waveform duration is short, and start-end events occur frequently. For the dishwasher, the overall model disaggregation fits well, and similarly, for the spikes in the waveform, the model does not fit well. A total of three troughs are decomposed in Fig. 5(e), and one trough is incorrectly decomposed. The analysis of the appliance disaggregation results shows that the model disaggregation is highly accurate, which proves that the RUCGAN model proposed in this paper has strong stability and robustness.

The U-Net network is used in the seq2subseq model, and we use the ResU-Net, so we summarize the excellent performance of the residual module on both datasets through experiments, and the results are shown in Fig. 6 and Fig. 7. Analyzing the performance of the models on the UK-DALE dataset, both models have an excellent disaggregation for the dishwasher, but the U-Net appears burred and, therefore, less stable than the residual model for the left-hand waveform. For the microwave oven, the U-Net showed a large amplitude at the top of both waveforms, while the ResU-Net remained more stable. For the kettle, due to its short running time, both models showed errors in predicting the start and end times, but the errors in the residual model were smaller, and the residual model remained stable at the peak in the kettle power disaggregation. For fridges, due to long working hours and frequent switching of operating states, U-Net has more errors in start-end time prediction and the waveform tears in the disaggregation process due to the instability of the model. Since the results of the Washingmachine are similar under both models therefore omitted. This shows that the residual network can mitigate the phenomenon of model gradient disappearance, and can optimize the network while the network deepens. The powerful local feature extraction ability of the residual module is used to obtain the deep features of the power sequences and guide the CGAN with powerful context awareness so that the before-and-after information of the power sequences can be well explored.

For the REFIT data, the RUCGAN model proposed in this paper still shows good disaggregation performance, as shown in Fig. 7. U-Net still exposes problems such as start-stop time disaggregation errors and waveform tearing in the REFIT dataset, and frequent un-disaggregation and mis-disaggregation problems occur in Fig. 7(b). The experiments on different datasets demonstrate both the effectiveness and the generalization of the ResU-Net module in RUCGAN.

C. COMPARISON OF EVALUATION METRICS AND DISAGGREGATION RESULTS

It is shown in Fig. 8 that the disaggregation performance of the RUCGAN model proposed in this paper is improved for all the five target appliances in the UK-DALE dataset. In terms of accuracy, washing machines, dishwashers and fridges have improved significantly, while kettles and microwaves have seen limited improvement as the accuracy of previous algorithms to decompose them has been close to 1. However, the proposed algorithm maintains the best performance compared to other algorithms in this paper. Recall and precision describe the classification of individual categories, which are not holistic and are intermediate quantities of the F-score, so attention is focused on the analysis of the F-score. Except for the fridge, which had limited improvement, the remaining four appliances improved significantly. We also found that previous algorithms performed poorly in disaggregation for appliances with complex operating states, especially for the Seq2point model. The disaggregation results of the F-score of the RUCGAN model all exceeded 70% and even exceeded 80%, except for the dishwasher.

The RUCGAN model proposed in this paper, and the Seq2subseq and VAE models all belong to the Encoder-Decoder framework model, and it can be seen from Table 5 that the Encoder-Decoder framework model outper-

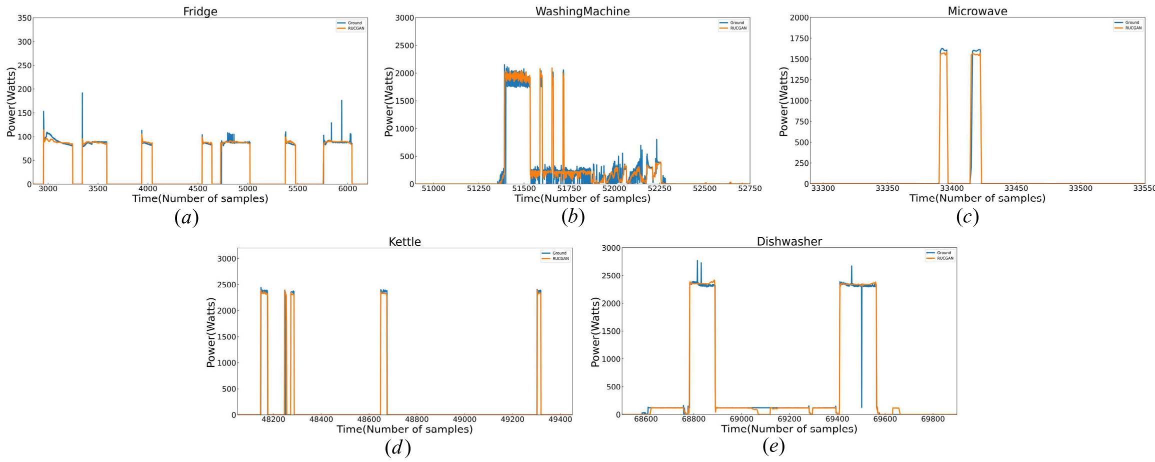

FIGURE 5. UK-DALE House 2: decomposing results of RUCGAN. (a) Fridge. (b) WashingMachine. (c) Microwave. (d) Kettle. (e) Dishwasher. The blue solid line represents the true values, and the orange solid line represents the decomposition values of the RUCGAN model.

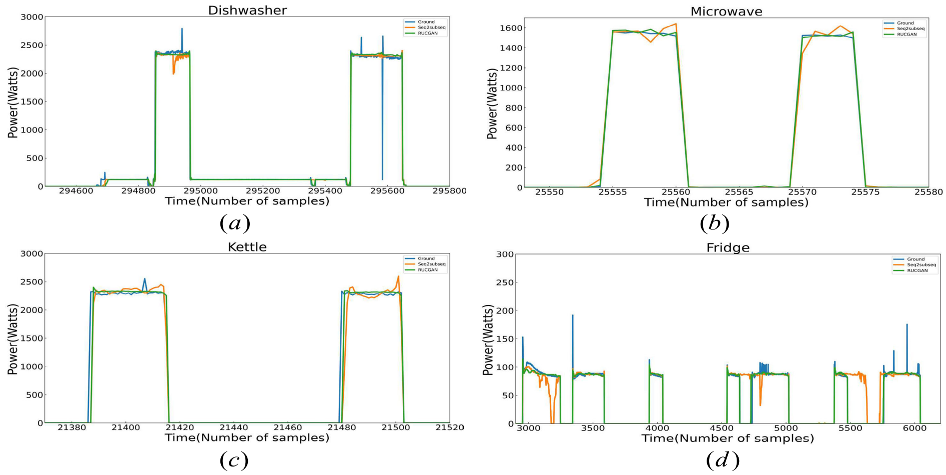

FIGURE 6. Positive effect of ResU-Net module on the model (UK-DALE). Only RUCGAN uses the ResU-Net module. (a) Dishwasher. (b) Microwave. (c) Kettle. (d) Fridge. The blue solid line represents the true values, the orange solid line represents the decomposition values of the comparison model Seq2subseq, and the green solid line represents the decomposition values of the RUCGAN model.

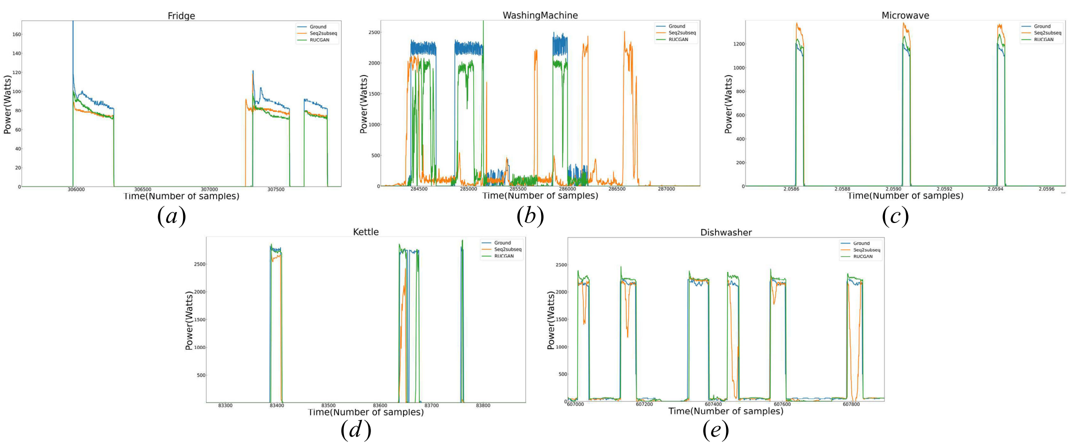

FIGURE 7. Positive effect of ResU-Net module on the model (REFIT dataset). (a) Fridge. (b) WashingMachine. (c) Microwave. (d) Kettle. (e) Dishwasher. The blue solid line represents the true values, the orange solid line represents the decomposition values of the comparison model Seq2subseq, and the green solid line represents the decomposition values of the RUCGAN model.

forms the other models in general. These models have good disaggregation performance for load appliances with simple structures and no complex operation state, for example, the disaggregation of kettle and microwave oven is better than

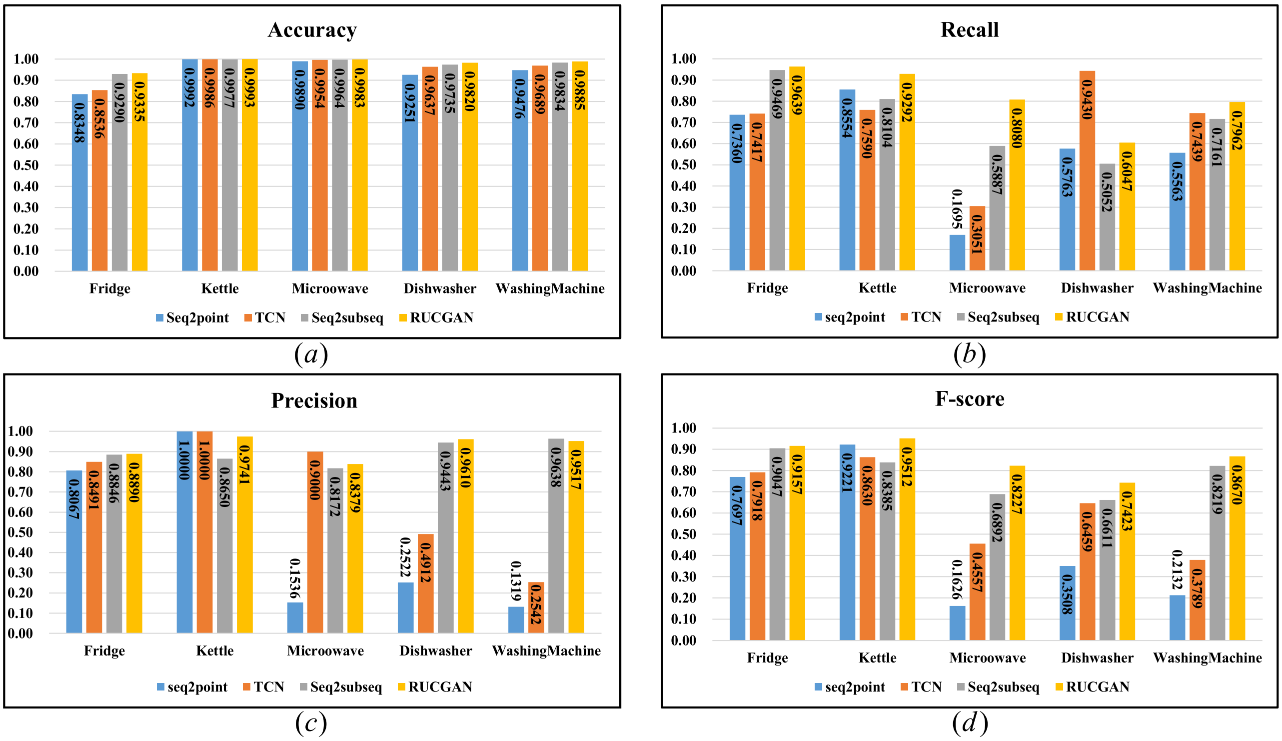

FIGURE 8. Load disaggregation classification results. (a) Accuracy. (b) Recall. (c) Precision. (d) F-score.

TABLE 5. The appliance-level mae(watt) for uk-dale data. best results are shown in bold and underlined.

Metric	Methods	WM	KE	MW	FR	DW	Overall
MAE	LSTM	109	16	20	36	168	69.8
	Seq2point	10.1	7.4	8.6	20.8	27.7	14.92
	Seq2subseq	7.11	3.59	3.14	11.86	13.52	7.84
	TCN	9.49	4.10	12.61	16.38	23.26	13.17
	LSTM+	12.27	7.21	8.46	22.35	20.95	14.25
	VAE	6.20	6.10	5.10	15.10	**11.60**	8.82
	ResNet+	6.21	—	5.67	12.31	18.25	10.61
	RUCGAN	**5.27**	**2.57**	**2.60**	**8.35**	12.38	**6.23**

TABLE 6. Number of samples used for load disaggregation.

	FR	KE	MW	WM	DW
Number of samples	47464	14376	11792	8276	4048

the other three appliances. The Seq2subseq and VAE models are only based on the Encoder-Decoder framework or the adoption of the U-Net, and from the data, the simple application of the U-Net does not improve the performance of the Encoder-Decoder framework model. But the performance of the RUCGAN model proposed in this paper is improved by 20.5% and 29.3% compared with Seq2subseq and VAE, respectively, proving that CGAN, U-Net and residual modules are deeply fused in this paper rather than simply stacked. The performance improvement of the RUCGAN model is more obvious than the model of the non-Encoder-Decoder frame work. Compared with LSTM, Seq2point, LSTM+, ResNet+, and TCN, the performance is improved by 91.1%, 58.2%, 52.7%, 41.2% and 39.2%, respectively. We note that for the power disaggregation of the dishwasher the results are mostly above 20 except

for the three models of the Encoder-Decoder framework. Table 6 shows the number of samples of each appliance in the

TABLE 7. The appliance-level mae(watt) for refit data. best results are shown in bold and underlined.

Metric	Methods	WM	KE	MW	FR	DW	Overall
	Seq2point	23.8	14.0	8.9	23.9	10.1	16.14
	Seq2subseq	20.7	6.53	6.48	20.59	9.37	12.73
MAE	LSTM+	13.21	11.56	8.33	19.24	13.41	13.10
	VAE	**<u>10.7</u>**	7.1	10.8	21.6	23.4	14.72
	RUCGAN	17.88	**<u>6.10</u>**	**<u>6.20</u>**	**<u>18.97</u>**	**<u>8.14</u>**	**<u>11.46</u>**

UK-DALE dataset used for training, with the dishwasher having the fewest number. This explains one of the reasons why the model of the Encoder-Decoder architecture outperforms other models; the framework model has data regeneration capability, and if the input is N samples, the network learns 2N samples. Therefore, we can appropriately increase the number of training samples to improve the disaggregation accuracy of the model.

In this paper, MAE metrics were also calculated and analyzed for five appliances in the REFIT dataset, as shown in Table 7. Since models such as ResNet+ use the UK-DALE and REDD datasets and the code is not open source, the remaining models were selected for the analysis of the MAE metrics comparison for the REFIT dataset. It can be seen that the MAE values of the proposed model in this paper decreased by 9.9%, 12.5%, 22.1%, and 28.9% compared to Seq2subseq, LSTM+, VAE, and Seq2point, respectively. For appliances with high power and simpler operation modes, such as microwave ovens, the MAE values of the above models are kept low. While for relatively complex appliances such as fridges and dishwashers, the deep feature extraction and context awareness of the RUCGAN model makes its power decomposition prediction much more accurate than other models. There were exceptions, however, where the washing machines in the VAE model showed better performance in terms of MAE metrics.

In summary, the model proposed in this paper uses the residual module with powerful local feature extraction capability and U-Net with depth perception capability to make the RUCGAN model perform its proper strength. The comparison with the Seq2subseq model shows the advantages of ResU-Net, and the comparison with the rest of the models reflects that generative adversarial network is one of the effective methods to solve the NILM problem. Tests on the UK-DALE and REFIT datasets show that the proposed model is generalizable.

VI. CONCLUSION

In this work, we proposed RUCGAN, a non-intrusive load decomposition algorithm based on the CGAN model, aiming to improve the model's decomposition accuracy and generalization performance. Firstly, we introduced the U-Net to process time series data, which preserves feature information and accelerates gradient propagation through skip connections. Secondly, we fused the residual block with the U-Net to reduce network parameters and simplify computation using residual connections, thereby avoiding the problem of model degradation. Our method achieved significant performance gains through the development of a specific model and the improvement of the encoding-decoding structure for time series data. Experimental results on different datasets showed that RUCGAN effectively reduced the MAE index and maintained high robustness performance. Furthermore, it also reduced the model size to some extent. In summary, compared to existing publicly available load decomposition algorithms, RUCGAN has significant advantages in terms of decomposition accuracy, model size, and stability. However, future research should focus on reducing the training time of the encoding-decoding structure model to achieve real-time decomposition.

REFERENCES

[1] D. Vine, L. Buys, and P. Morris, "The effectiveness of energy feedback for conservation and peak demand: A literature review," *Open J. Energy Efficiency*, vol. 2, no. 1, pp. 7–15, 2013, doi: 10.4236/ojee.2013.21002.

[2] F. B. Gurbuz, R. Bayindir, and H. I. Bulbul, "A brief review of non-intrusive load monitoring and its impact on social life," in *Proc. 9th Int. Conf. Smart Grid (icSmartGrid)*, Setúbal, Portugal, Jun./Jul. 2021, pp. 289–294.

[3] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992, doi: 10.1109/5.192069.

[4] T. Zia, D. Bruckner, and A. Zaidi, "A hidden Markov model based procedure for identifying household electric loads," in *Proc. 37th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Melbourne, VIC, Australia, Nov. 2011, pp. 3218–3223.

[5] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. SIAM Int. Conf. Data Mining*, Mesa, AZ, USA, Apr. 2011, pp. 747–758.

[6] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, pp. 673–701, Feb. 2013.

[7] A. Zoha, A. Gluhak, M. Nati, and M. A. Imran, "Low-power appliance monitoring using factorial hidden Markov models," in *Proc. IEEE 8th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, Melbourne, VIC, Australia, Apr. 2013, pp. 527–532, doi: 10.1109/ISSNIP.2013.6529845.

[8] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proc. 26th Conf. Artif. Intell. (AAAI)*, Toronto, ON, Canada, Jul. 2012, pp. 356–362.

[9] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. Artif. Intell. Statist.*, La Palma, Spain, Apr. 2012, pp. 1472–1482.

[10] M.-S. Tsai and Y.-H. Lin, "Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation," *Appl. Energy*, vol. 96, pp. 55–73, Aug. 2012.

[11] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of VsI trajectory based load signatures for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 870–878, Mar. 2014.

[12] T. Saitoh, T. Osaki, R. Konishi, and K. Sugahara, "Current sensor based home appliance and state of appliance recognition," *SICE J. Control, Meas., Syst. Integr.*, vol. 3, no. 2, pp. 86–93, Mar. 2010.

[13] Y. Lin and M. Tsai, "Non-intrusive load monitoring by novel neuro-fuzzy classification considering uncertainties," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2376–2384, Sep. 2014.

[14] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "PALDi: Online load disaggregation via particle filtering," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 467–477, Feb. 2015, doi: 10.1109/TIM.2014.2344373.

[15] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 448–461, Jan. 2019.

[16] H. Xu, C. Li, M. M. Rahaman, Y. Yao, Z. Li, J. Zhang, F. Kulwa, X. Zhao, S. Qi, and Y. Teng, "An enhanced framework of generative adversarial networks (EF-GANs) for environmental microorganism image augmentation with limited rotation-invariant training data," *IEEE Access*, vol. 8, pp. 187455–187469, 2020.

[17] T. Sun, "End-to-end speech emotion recognition with gender information," *IEEE Access*, vol. 8, pp. 152423–152438, 2020.

[18] D. Wang, J. Su, and H. Yu, "Feature extraction and analysis of natural language processing for deep learning English language," *IEEE Access*, vol. 8, pp. 46335–46345, 2020.

[19] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, Seoul, South Korea, Nov. 2015, pp. 55–64.

[20] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, "Denoising autoencoders for non-intrusive load monitoring: Improvements and comparative evaluation," *Energy Buildings*, vol. 158, pp. 1461–1474, Jan. 2018.

[21] A. Langevin, M.-A. Carbonneau, M. Cheriet, and G. Gagnon, "Energy disaggregation using variational autoencoders," *Energy Buildings*, vol. 254, Jan. 2022, Art. no. 111623, doi: 10.1016/j.enbuild.2021.111623.

[22] V. Piccialli and A. Sudoso, "Improving non-intrusive load disaggregation through an attention-based deep neural network. Energies 2021, 14, 847," in *Energy Data Analytics for Smart Meter Data*, 2021, p. 63.

[23] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, vol. 32, no. 1, pp. 2604–2611.

[24] W. Yang, C. Pang, J. Huang, and X. Zeng, "Sequence-to-point learning based on temporal convolutional networks for nonintrusive load monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[25] Y. Pan, K. Liu, Z. Shen, X. Cai, and Z. Jia, "Sequence-to-subsequence learning with conditional gan for power disaggregation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 3202–3206.

[26] J. Song, H. Wang, M. Du, L. Peng, S. Zhang, and G. Xu, "Non-intrusive load identification method based on improved long short term memory network," *Energies*, vol. 14, no. 3, p. 684, Jan. 2021.

[27] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[28] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[32] Z. Nie, Y. Yang, and Q. Xu, "An ensemble-policy non-intrusive load monitoring technique based entirely on deep feature-guided attention mechanism," *Energy Buildings*, vol. 273, Oct. 2022, Art. no. 112356.

[33] J. Kelly and W. Knottenbelt, "The U.K.-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five U.K. Homes," *Sci. Data*, vol. 2, no. 1, pp. 1–14, Mar. 2015.

[34] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study," *Sci. Data*, vol. 4, no. 1, pp. 1–12, Jan. 2017.

[35] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proc. 5th Int. Conf. Future Energy Syst.*, Cambridge, U.K., Jun. 2014, pp. 265–276.

[36] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 6, p. e1265, 2018.

JINLONG WANG (Graduate Student Member, IEEE) was born in 1998. He received the B.E. degree in electronic information engineering from the Tongda College, Nanjing University of Posts and Telecommunications, Yangzhou, China, in 2020. He is currently pursuing the M.E. degree in information and communication engineering with the Shanghai University of Electric Power (SUEP). His research interest includes non-intrusive load monitoring (NILM).

CHENGXIN PANG (Member, IEEE) received the Doctor degree from the Shanghai University of Electric Power (SUEP). He is currently a Shanghai High-Level Distinguished Expert and a Professor with SUEP. His research topics have been firstly magneto-optics isolator in IEF (Paris-Sud University, Orsay, France) and hybrid silicon photonics in the Orange Laboratories, France, and in LCF (IOGS, Palaiseau, France). Since 2015, he has been involved in the research power Internet of Things, NILM, intelligent information perception, and machine vision.

XINHUA ZENG (Member, IEEE) received the Doctor degree. He is currently the Deputy Director of the Network and System Center, Institute of Technology, Fudan University, and the Director of the Center of Hefei Institute of Technological Innovation, Chinese Academy of Sciences. His current research interests include biological information collection, intelligent information perception, processing, visual image, and deep learning.

YONGBO CHEN is currently a Senior Engineer and a Chief Engineer with ZTE Power Industry. He engaged in smart grid and ICT related fields for more than 20 years. In ZTE, he is also responsible for the research, product planning and solution development of power digitization, 5G application, and the IoT.