energies

MDPI

Article

Non-Intrusive Load Monitoring Based on Swin-Transformer with Adaptive Scaling Recurrence Plot

Yongtao Shi, Xiaodong Zhao, Fan Zhang * and Yaguang Kong *

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
* Correspondence: zhangfan@hdu.edu.cn (F.Z.); ygkong@hdu.edu.cn (Y.K.)

Abstract: Non-Intrusive Load Monitoring (NILM) is an effective energy consumption analysis technology, which just requires voltage and current signals on the user bus. This non-invasive monitoring approach can clarify the working state of multiple loads in the building with fewer sensing devices, thus reducing the cost of energy consumption monitoring. In this paper, an NILM method combining adaptive Recurrence Plot (RP) feature extraction and deep-learning-based image recognition is proposed. Firstly, the time-series signal of current is transformed into a threshold-free RP in phase space to obtain the image features. The Euclidean norm in threshold-free RP is scaled exponentially according to the voltage and current correlation to reflect the working characteristics of different loads adaptively. Afterwards, the obtained adaptive RP features can be mapped into images using the corresponding pixel value. In the load identification stage, an advanced computer vision deep network, Hierarchical Vision Transformer using Shifted Windows (Swin-Transformer), is applied to identify the adaptive RP images. The proposed solution is extensively verified by four real, measured load signal datasets, including industrial and household power situations, covering single-phase and three-phase electrical signals. The numerical results demonstrate that the proposed NILM method based on the adaptive RP can effectively improve the accuracy of load detection.

Keywords: Non-Intrusive Load Monitoring; adaptive scaling; Recurrence Plot; Swin-Transformer multi-head attention

Citation: Shi, Y.; Zhao, X.; Zhang, F.; Kong, Y. Non-Intrusive Load Monitoring Based on Swin-Transformer with Adaptive Scaling Recurrence Plot. *Energies* **2022**, *15*, 7800. https://doi.org/10.3390/en15207800

Academic Editor: Adolfo Dannier

Received: 23 September 2022
Accepted: 18 October 2022
Published: 21 October 2022

1. Introduction

Non-Intrusive Load Monitoring (NILM) technology realizes building energy monitoring and energy consumption analysis with as few physical sensing devices as feasible, hence minimizing monitoring expenses and user disruption [1]. Effective implementation of NILM will benefit significantly in all respects [2]. By knowing which appliances consume more energy, consumers can better understand their energy consumption; so, they can make smarter decisions about their electrical habits. Most consumers are unaware of the energy they consume and their impact on the environment. Increasing awareness of energy efficiency can help them use appliances more efficiently [3]. Some studies show that providing consumers with detailed energy consumption data can reduce building energy consumption by 12% [4]. Further, feedback on energy use can be enriched with a specific set of recommendations that give consumers the opportunity to achieve specific measurable energy efficiency goals. It can even diagnose the use of household appliances remotely to inform consumers of unusual usage patterns of existing appliances.

Grid companies can also benefit from NILM. By strengthening their contact with customers, they can better understand the use and consumption of electricity. By exploring the power consumption behavior of customers, grid companies can effectively achieve market segmentation—that is, identifying customer clusters with similar needs. In this way, they can design more effective marketing strategies and even provide more choices when offering customized services. Understanding how customers consume electricity is critical for grid companies. Analyzing historical data can reveal the operating mode of each

power device and allow grid companies to provide more personalized recommendations to meet the needs of each customer. Moreover, it can also achieve more accurate day-ahead or short-term load forecasting, which promotes the implementation of demand response strategy [5,6]. Demand response means that consumers can be motivated to limit or time transfer the use of certain electrical appliances, giving grid operators the opportunity to create a more accurate match between power supply and demand. Therefore, the energy-saving potential resulting from NILM is expected to have a positive impact on the environment. In addition, NILM systems can diagnose failures or abnormal operations in a non-invasive manner, and timely alerts can avoid potentially catastrophic failures in buildings [7].

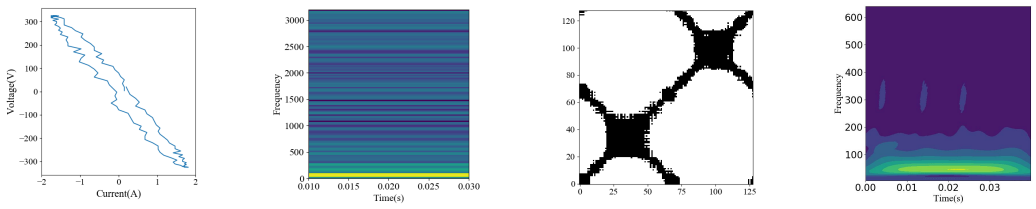
NILM was firstly proposed by Dr. Hart in 1992 [8]; it has been developed extensively over the last decade due to the rapid growth of machine learning and deep learning. Numerous experts have conducted more extensive research on it [1]. A typical NILM technology system is usually composed of four parts: data acquisition, event detection, feature extraction, and load decomposition [9].

Data acquisition involves collecting electrical signal from the bus, which demands high efficiency, high precision, and strong anti-interference capability. Then, the event detection can be used to accurately identify various startup–shutdown and abnormal change events. After obtaining the load events, feature extraction is utilized to extract or convert the high-dimensional feature inside the load electrical signal to filter out the interference information and highlight the distinctions between different types of loads. Afterwards, load identification uses pattern matching, machine learning, deep learning, and other methods to classify the extracted load features. To make the load energy consumption evident, power decomposition splits the power consumption curve based on the outcome of load identification. As the power environment becomes more and more complex and the number of loads increases, new requirements are put forward for the universality of NILM technology. By adjusting the parameters adaptively and adapting various power environments, NILM technology can be widely used.

In NILM, feature extraction and load identification are crucial. Excellent feature extraction and accurate load identification can effectively improve the power decomposition efficiency, so as to grasp more details of building energy consumption. Numerous researchers have proposed diverse feature transformation and identification strategies for the acquired electrical signals. Converting electrical signals into images and load recognition into image recognition has become a popular research direction in the field of NILM in recent years due to the rapid development of image recognition [1].

To improve the success rate of the image recognition algorithm, the original electrical signals are converted into image features. The graphs obtained by taking the current (I) and voltage (V) as the horizontal and vertical axes, respectively, are called V-I trajectory, which are commonly used as time-domain load features, as shown in Figure 1a. V-I trajectory is often used as a feature [10] on which weight information [11] or discrete color-coded background [12] can be added to generate a new load feature. Others clip V-I trajectory templates and simplify image pyramids [13]. A multilevel load classification algorithm is designed, and the validity of the algorithm is verified on the actual dataset. In addition, the symmetry, area, and peak [14] of the V-I trajectory are also used to explore deeper and more comprehensive load features.

In practice, V-I trajectory is easy to obtain. It can accurately depict the time-domain difference between different loads; so, it usually works well in scenarios where the environment is stable, there are few kinds of loads, and there are abundant sample data. However, it loses the frequency domain characteristics. In addition, it is not favorable to the deep information mining of image recognition algorithms since its feature information is very concentrated and singular. Therefore, the environmental adaptability of V-I trajectory is not high, and it is difficult to efficiently use the load identification algorithm, which leads to certain limitations in the application scope of NILM technology.



(**a**) (**b**) (**c**) (**d**)

Figure 1. Different types of load map features. (**a**) V-I trajectory. (**b**) Spectrogram. (**c**) Recurrence plot. (**d**) Continue wavelet transform.

In addition to time-domain data, some researchers used Fourier transform to convert electrical signals into frequency-domain features, such as fast Fourier spectrogram [15], as shown in Figure 1b, for load recognition. Although the fast Fourier transform may accurately characterize the frequency characteristics of a signal, it will lose a portion of the information in the time domain, preventing us from observing the frequency shift with time. Considering time–frequency domain analysis, [16] adopted Continuous Wavelet Transform (CWT) to extract load features, shown in Figure 1d. CWT can localize and uniformly present information in the time–frequency domain at the same time, but its internal feature distribution location is concentrated, which can easily cause feature confusion among loads. To maximize the use of high-dimensional features of electrical signals in the time–frequency domain, some researchers have fused the features of various time–frequency domain graphs (Markov transition field, Gramian angular field, spectrogram, etc.) obtained from electrical signal conversion based on the principle of image four channel synthesis to combine the advantages of time–frequency domain features [15]. Multi-type image fusion can successfully utilize the multidimensional depth characteristics of electrical signals. Relatively, it will invariably result in information coverage, feature loss, or redundancy.

Introducing frequency domain information can improve the diversity of feature extraction in NILM technology; whether it is pure frequency domain analysis, time–frequency domain analysis, or multiple types of time–frequency domain feature superposition, it can effectively improve the accuracy and applicability of NILM technology. However, frequency feature acquisition requirements and calculation complexity are high, and load identification is prone to a certain time lag, which is not suitable for scenarios with high timeliness requirements and frequent load events. In addition, the information distribution differentiation of time–frequency characteristics corresponding to different loads is small. So, it depends on mature load identification algorithms in practical applications.

Besides, some researchers employed a non-linear analytic technique to identify load features. As illustrated in Figure 1c, the Recurrence Plot (RP) analysis in phase space is employed to deal with the non-linear load current [17]. Then, ref. [18] overlaid the weight information to create the weighted recurrence plot functionality. Recurrence plots show the time–frequency, non-linearity, and periodicity of electrical signals. Besides, its visual feature information is abundant and evenly distributed. Compared with other load characteristics, it can maximize the benefits of the image recognition method. However, the stability and universality of RP are insufficient since the hyper-parameter changes have a significant effect on the recognition outcomes and depend on the experimental precision. Inspired by [19], we offer an adaptive signal image feature that integrates RP with difference of V and I. It effectively utilizes the benefits of the recurrence plot feature and reduces the computational complexity of hyper-parameters. Therefore, it can reduce the dependence of the NILM system on hyper-parameters. Adaptive parameter adjustment can have stronger environmental adaptability in actual use and reduce later maintenance cost.

Once the load image feature dataset has been gathered, an identification algorithm can be constructed to learn and identify existing data. Common techniques for recognition can be classified as optimum combinations, machine learning, and deep learning [20].

Optimal combination seeks to match observed power measurements with all possible device power signal combinations [21]. In [22], the Particle Swarm Optimization (PSO) method was upgraded by introducing temporal probability and integrated with many optimal combination algorithms as the recognition classifier of NILM to obtain good classification results. When the number of loads is minimal and their properties are known, these methods have a good decomposition effect and are intuitive in theory. Load identification is, however, a non-deterministic polynomial hard (NP-hard) task, making it difficult to enhance optimization efficiency. Additionally, this method assumes that the characteristics of each load satisfy the superposition, but not all characteristics satisfy this criterion. Furthermore, this approach requires a comprehensive feature library of the load to perform accurate recognition, which is typically difficult to obtain in actual applications.

Typical machine learning methods such as decision tree and random forest are frequently employed as the primary body of the load identification module [23]. In addition, to improve the accuracy of load identification, some researchers used the fundamental amplitude and harmonic amplitude of the collected electrical signal as load characteristics and a support vector machine as a load identification model to construct a framework for load identification [24]. Another study developed a random forest ensemble classifier for the categorization of unbalanced loads [9]. The method of machine learning can increase the accuracy of load identification, but its scalability is limited. As the number of load categories increases, the accuracy of identifying similar loads will be compromised.

Deep learning enables computers to learn and recognize independently by deeply mining various load features. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory Networks (LSTM), etc. are widely used in NILM technology. For example, ref. [25] used CNN to perform multi-class device identification studies on transient electrical signals sampled at higher frequencies. It is observed that the fine-tuned deep architecture outperforms the traditional multi-class classification algorithms. Similarly, ref. [26] proposed concatenated CNN to remove the influence of noises, voltage fluctuations, and background loads from the target load signals for effective load recognition in a shorter duration. The reported results outperformed traditional NILM techniques. In addition, ref. [27] introduced a novel denoising auto-encoder to the NILM field, which has garnered much interest in recent years. Compared with the commonly used NILM method, the F1 score was increased by more than 11%. Moreover, ref. [28] used the ability of deep learning network LSTM to store key information in memory to identify the high-dimensional features obtained from Sequence-to-point learning. As a novel recognition technique, deep learning possesses significant feature mining and differential memory capabilities, hence enhancing the accuracy of NILM.

With the introduction of edge cloud gateways and 5G technology, high computing power and massive data volume needs of deep learning no longer impact the cost of NILM equipment. The NILM field is becoming appreciative of its excellent recognition rate and scalability. In this paper, the newly emerging Swin-Transformer network is used as identification algorithm, and the load identification framework is constructed in conjunction with adaptive variable parameter features. The approach has broad applicability and is applicable to both single-phase and three-phase electrical signals. The fault tolerance rate of NILM can be improved by introducing deep learning method. In addition, by adapting single-phase to three-phase electrical signals, the proposed method can be applied to many power consumption scenarios such as home and industry.

In this paper, an NILM method combining adaptive adjustment and deep learning is proposed. The correlation of voltage and current is combined with the threshold-free RP in phase space to figure out a method of feature extraction adaptively, aiming to address the issues of information centralization, simple shape, and difficult formulation of hyper-parameters in common load image features. The load feature images are mapped, which can be recognized by means of the advanced Swin-Transformer with multi-head self-attention and shifted window mechanism. In summary, the major technological innovations are as follows:

(1) An adaptive scaling RP (ASRP) is realized by transforming the current signal into a threshold-free RP in phase space and scaling it exponentially according to the correlation between voltage and current, which effectively reflects the working characteristics of different loads.
(2) Swin-Transformer is utilized to efficiently characterize the high-dimensional latent information in adaptive RP. In addition, the shifted window mechanism in the network can lower the computational complexity of the model. Eventually, efficient and accurate non-intrusive load identification is attained.
(3) Four measured load signal datasets, covering industrial and domestic electricity scenarios, including single-phase and three-phase electrical signals, are utilized to verify the generalizability of the proposed method.

The subsequent content is organized as follows: Section 2 covers the load identification methods employed in detail, Section 3 describes the specific experimental design concepts, Section 4 demonstrates the relevant experimental results, and Section 5 summarizes the methodologies and contributions.

2. Load Identification Based on Adaptive Scaling Recurrence Plot

The electric signal data V_o and I_o with start–stop events are extracted by collecting the electric signal data in the user bus and performing load event detection. Then, they undergo preprocessing, feature extraction, and load identification. Initially, this section performs the preprocessing of the electrical signal by obtaining the current difference I_k and voltage V_k before and after each start–stop event from I_o and V_o. Then, the ASRP is transformed, and the scale of the transformed ASRP is determined according to the phase number of the electrical signal. Finally, the deep learning network is used to train and identify the load feature dataset. Later, the trained load identification model will be deployed to a local area to realize the online identification of load and visualize the building energy use through power decomposition module. The overall route of the method is shown in Figure 2.

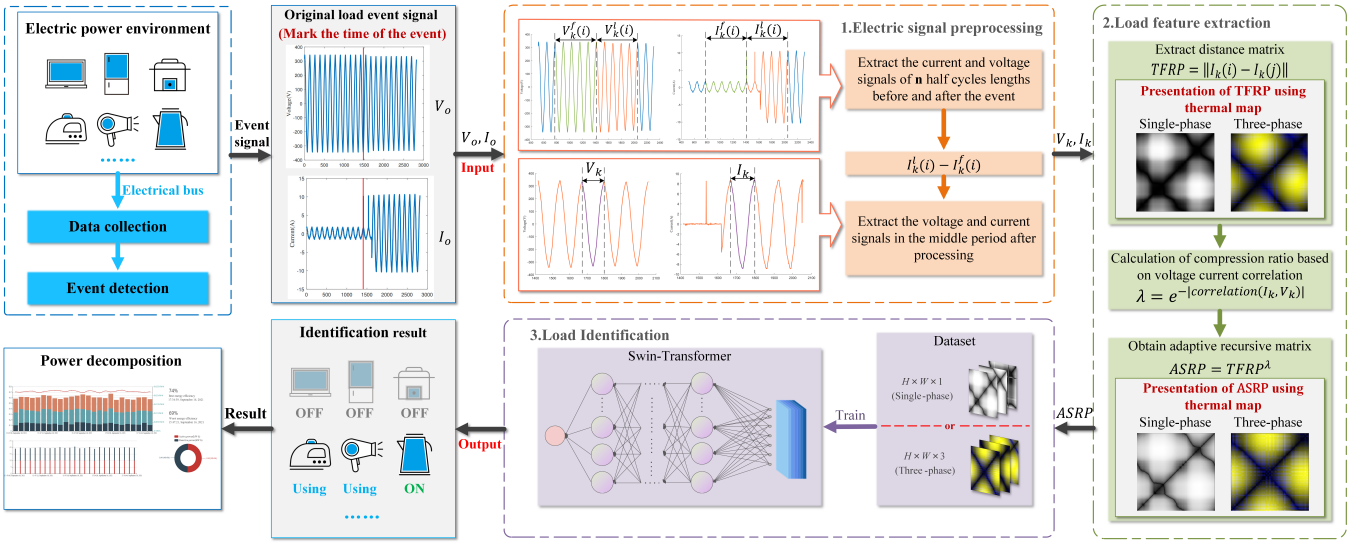


Figure 2. Technology roadmap for non-intrusive load monitoring based on adaptive scaling recurrence plot.

2.1. Electric Signal Preprocessing

Since there may be one or more loads in use in the acquired original electric signals V_o and I_o, and a certain type of load performs the start–stop operation at the current time, it is necessary to extract the load electric signal data and perform the start–stop operation at the current time based on them. Benefiting from the excellent superposition of current data and stability of voltage data, the difference in the current data before and after start–stop and

the voltage data after start–stop can be used as the load electrical signal data of start–stop operation at the current moment [29]. The overall procedure is depicted in Figure 3.

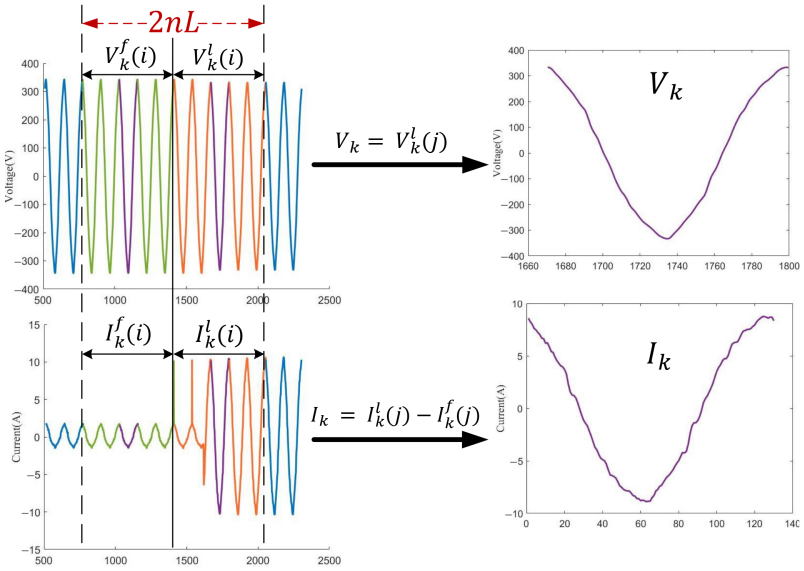


Figure 3. Load electrical signal data preprocessing process.

Periodic calibration is required when calculating the current data difference in the instantaneous electric signal. In practice, it is difficult to observe the periodicity of the current curve, whereas the periodicity of the voltage curve is consistent and evident. Consequently, using the voltage curve as the standard, the current and voltage of n cycles lengths prior to and following the start–stop instant are obtained from V_o and I_o, respectively. With the accuracy of the event detection method, n can be modified. n can be a tiny integer value $[2, 5]$ if the technique for event detection is highly accurate. If the algorithm for event detection has low precision, n can have a bigger integer value $[4, 10]$. n is 5 in this article.

Assume that the current and voltage sequences before start–stop are $I_k^f(i)$, $V_k^f(i)$, and the current and voltage sequences after start–stop are $I_k^l(i)$, $V_k^l(i)$, where $k \in \{1, 2, 3\}$, represents the electrical signal data of the k phase. $i \in \{1, 2, 3, \ldots, nL\}$ represents the i value in the sequence. L is the single-cycle length of the electrical signal, with the calculation formula being $L = f_s/f$, where f_s is the sampling frequency and f is the power supply frequency. After that, the difference in current before and after start–stop in the middle cycle is taken as the processed current feature I_k, and the voltage is taken as the processed voltage feature V_k. The calculation method is as follows:

$$\begin{cases} I_k = I_k^l(j + \lfloor \frac{nL}{2} \rfloor) - I_k^f(j + \lfloor \frac{nL}{2} \rfloor) \\ V_k = V_k^l(j + \lfloor \frac{nL}{2} \rfloor) \end{cases}. \quad j = 1, 2, ..., L \tag{1}$$

where $\lfloor . \rfloor$ means rounding down.

2.2. Load Feature Extraction

After the load electric signal data (I_k and V_k), which perform start–stop operation at the current time, are obtained through data preprocessing, feature extraction can be performed. It can amplify the feature difference between various loads, thus effectively improving the accuracy of load identification algorithm and the analysis efficiency of NILM. RP is a recurrence visualization tool for measuring phase space state sequences [30]. RP can visualize the changing trend of a non-linear system. It can explain the internal structure of time series and give prior knowledge about similarity, amount of information, and predictability. Besides, it is considered a representation of the global correlation structure of a system, as it can reflect the autocorrelation properties of the system under study at all

possible time scales. Using the obtained I_k sequence as an illustration, the following matrix depicts RP:

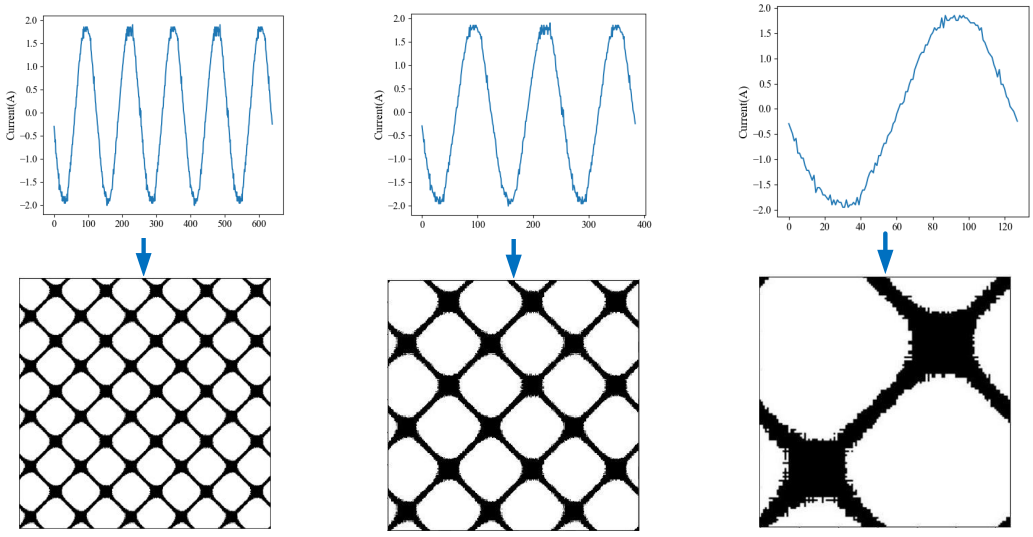
$$R_{i,j,k}(I_k, \varepsilon) = \Theta(\varepsilon - \| I_k(i) - I_k(j) \|) \quad i, j = 1, 2, ..., L_r \tag{2}$$

where L_r is the sequence length, which is the length of the extracted single-period electrical signal sequence; ε is the distance threshold at which recursion occurs; and $\Theta(x)$ is the Heaviside function.

$$\Theta(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{3}$$

In Equation (2), $\| \cdot \|$ means norm calculation. $L1$ norm, $L2$ norm (Euclidean norm), and $L\infty$ norm are generally used. In this article, we choose to use the $L2$ norm. RP mainly draws binary recurrence matrices in black and white colors. Among them, if $R_{i,j,k} = 1$, then the coordinate point (i, j) corresponding to the k scale is black; if $R_{i,j,k} = 0$, then the coordinate point (i, j) corresponding to the k scale is white. According to the definition of matrix R, RP always has a black main diagonal, and matrix R is a symmetric matrix with respect to the main diagonal. RP is a powerful graphical technique that can reveal the complexity and non-stationarity of time series, as well as reveal more detailed information hidden in scalar time series.

RP is plotted based on load signals of different lengths obtained from actual sampling, see Figure 4. In RP analysis, longer signal input means more macroscopic phase space recursion, and microscopic differences are easily ignored. Since the load signal is a standard periodic signal, the corresponding RP will also show obvious periodicity, as in Figure 4a,b. The vertical distance between diagonal lines in RP is the cycle length. It can be seen in Figure 4c that single-cycle RP can present more subtle information. In this study, single-cycle signals are used to analyze so as to eliminate the repeated information of features, amplify the microscopic differences of different loads, and effectively extract the irregular mutation behavior in load signals.



(**a**) (**b**) (**c**)

Figure 4. Current sequence with different time length and corresponding RP. (**a**) Five cycles. (**b**) Three cycles. (**c**) Single cycle.

In the field of NILM, the reason why time series electrical signals are converted to RP features is that RP can reasonably expand the original information and enhance the visualization of features, shown in Figure 5a. Compared with the original time series signal, RP can make efficient use of the advanced graph recognition network to improve the load identification effect.

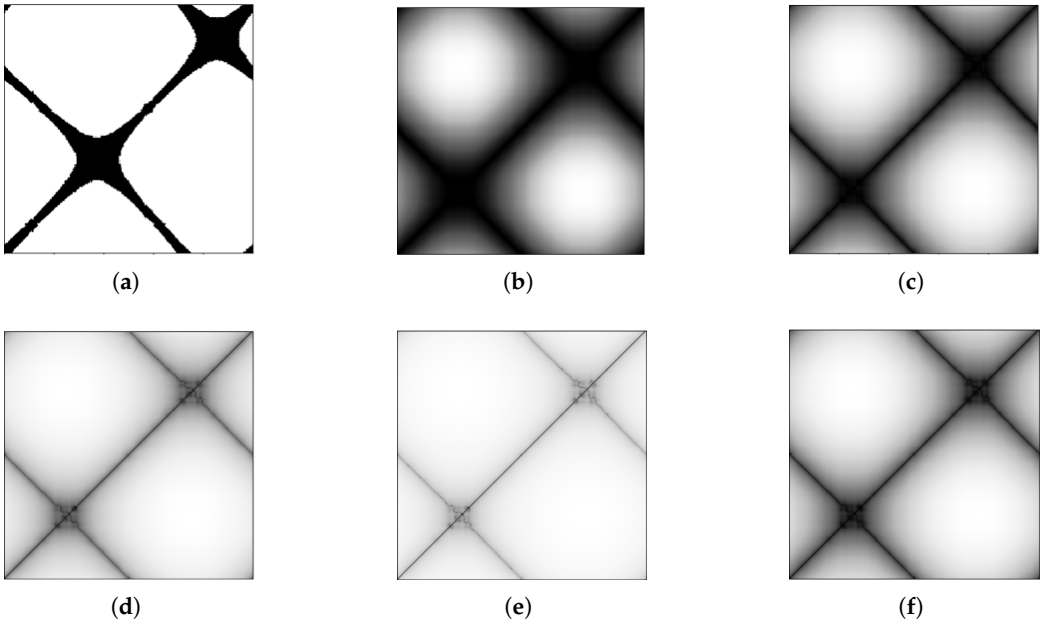


Figure 5. RP, SRP, and ASRP are visualized by heat maps (where (**b**–**e**) are the different effects of different hyper-parameter choices corresponding to SRP). (**a**) RP. (**b**) SRP ($\lambda = 1$). (**c**) SRP ($\lambda = 0.6$). (**d**) SRP ($\lambda = 0.2$). (**e**) SRP ($\lambda = 0.1$). (**f**) ASRP.

As RP has only binary values, the eigenvalues are too monotonous, and due to the Heaviside function, it will lose much information when it is applied to the NILM field, which is not conducive to the deep feature mining of image recognition algorithms. To highlight the main features without reducing the tiny features, the threshold-free RP (TFRP), which means RP that lacks the Heaviside function, is used as the load feature, shown in Figure 5b. The calculation formula is as follows:

$$TFRP_{i,j,k}(I_k) = \| I_k(i) - I_k(j) \| \quad i, j = 1, 2, ..., L. \tag{4}$$

TFRP can use multivariate values to represent a load. However, due to the lack of Heaviside function, the presented features are not clear enough, and too-loose a feature distribution is not conducive to load identification. To this end, we scale the TFRP exponentially by introducing scale ratio λ to obtain the Scaling Recurrence Plot (SRP). By comparing Figure 5b,d, after focusing TFRP features with λ, some details can be presented more obviously so as to improve the identification of different loads. SRP can be calculated by

$$SRP_{i,j,k}(I_k, \lambda) = \| I_k(i) - I_k(j) \|^{\lambda} \quad i, j = 1, 2, ..., L. \tag{5}$$

Whereas, the hyper-parameter λ has an important influence on the information presentation degree of SRP. If λ is too large, it is not conducive to highlighting detailed features, and if λ is too small, a large amount of information will be lost, as shown in Figure 5b,e. In addition, the phase difference information of current and voltage in electrical signals also plays an important role in load discrimination [19]. To solve the problem that the hyper-parameter λ of SRP has a great influence on the recognition results and incorporating phase difference information between current and voltage to improve load characteristic spacing, the correlation between I_k and V_k, $cor(I_k, V_k)$, is calculated to achieve adaptive adjustment of parameters. $cor(X, Y)$ means to calculate the correlation coefficient of the X and Y series. The commonly used correlation coefficients are $cor_{Pearson}$, $cor_{Spearman}$, and $cor_{Kendall}$. This paper uses the Pearson correlation coefficient. It can add new electrical signal features to the SRP, and improve the effective use of electrical signal data and the feature discrimination of different types of loads. In addition, to avoid information loss caused

by excessive compression, the range of compression is limited by the e index. The ASRP is obtained as follows:

$$\begin{aligned} \lambda_k &= e^{-|cor(I_k, V_k)|} \\ ASRP_{i,j,k}(I_k, V_k) &= \| I_k(i) - I_k(j) \|^{\lambda_k}, i, j = 1, 2, ..., L. \end{aligned} \tag{6}$$

The pseudocode of the above feature extraction process is summarized as follows (Algorithm 1):

Algorithm 1 Load feature extraction based on ASRP

Input: $I_k^d, V_k^d \quad (d = 1, 2, ..., D)$, I_k^d, V_k^d is current and voltage dataset after preprocessing, D is the total number of dataset.

repeat

$d \leftarrow d + 1$

Update $cor(I_k^d, V_k^d)$

Update $\lambda_k = e^{-|cor(I_k^d, V_k^d)|}$

Update $ASRP_{i,j,k}^d = \| I_k^d(i) - I_k^d(j) \|^{\lambda_k}$

until $d = D$

Output: $ASRP_{i,j,k}^d(i, j = 1, 2, ..., L; d = 1, 2, ..., D)$, k represents the electrical signal of the k-th phase in the dataset (single-phase: $k = 1$; three-phase: $k = 1, 2, 3$).

SRP and ASRP can be visualized using heat maps, as shown in the Figure 5. It can be seen from the image comparison that with the decrease in λ, the main features presented by the SRP will be closer to RP; however, on this basis, the surrounding subtle features will be retained. In addition, due to the limitation of the e index, the SRP will be compressed within a reasonable interval with the similarity of the V-I curves of different loads without loss of features due to excessive compression, which improves the feature distinction of loads with similar current curves.

2.3. Load Identification Based on Swin-Transformer

In order to efficiently identify the extracted load features, the advanced deep visual processing network Swin-Transformer is selected for load identification. Swin-Transformer [31] introduces a hierarchical structure and uses local computational self-attention within non-overlapping windows to reduce computational complexity to a linear relationship between image size. It effectively solves the drawbacks of single feature, high computational complexity, and long training time in the original transformer [32] network architecture. This maintains the deep mining characteristics of the original architecture, effectively reduces the training time, and adapts to multi-scale features.

Swin-Transformer proposes W-MSA (Windows Multi-head Self-Attention) and SW-MSA (Shifted Windows Multi-Head Self-Attention)—which are used in pairs, as shown in Figure 6—combined with LN (Layer Norm) and MLP (Multilayer Perceptron) together to form a Swin-Transformer Block. The computational flow is as follows:

$$\begin{aligned} \hat{z}^l &= \text{W-MSA}(\text{LN}(z^{l-1})) + z^{l-1} \\ z^l &= \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l \\ \hat{z}^{l+1} &= \text{SW-MSA}(\text{LN}(z^l)) + z^l \\ z^{l+1} &= \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1} \end{aligned} \tag{7}$$

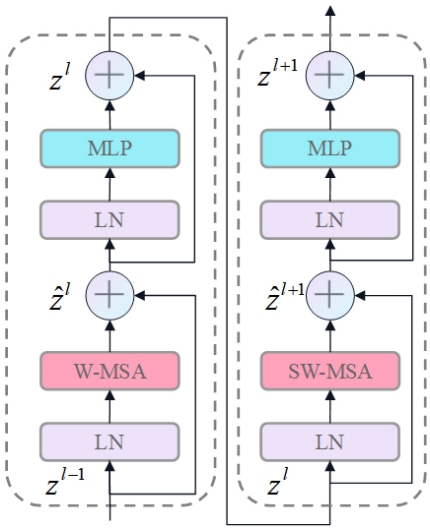


Figure 6. Swin-Transformer block structure.

MSA is a global self-attention calculation for input features [32], see Figure 7a. W-MSA effectively reduces computational effort by dividing the original features into $M \times M$ sized windows and performing MSA calculations on each individual window, as presented in Figure 7b. In addition, in order not to lose interactive information between windows, the original features are divided and transferred as in Figure 7c, and W-MSA is used again for calculation, a step called SW-MSA. Overall, the paired use of W-MSA and SW-MSA can effectively reduce the computational effort while improving the utilization of the original features.

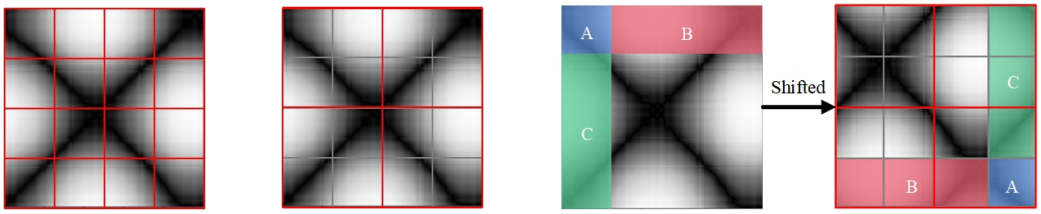


Figure 7. MSA, W-MSA, and SW-MSA calculation window selection difference display. (**a**) MSA. (**b**) W-MSA ($M = 2$). (**c**) SW-MSA($M = 2$): Select A, B and C areas in the original image as shown in the left figure, and then perform the transfer as shown in the right figure before window segmentation and calculation.

The Swin-Transformer network architecture used is shown in Figure 8. Since there are not many load categories for existing datasets in the NILM domain, the lightweight architecture Swin-T [31] is used as the principal part for identifying networks. The ASRP converted in the previous section is used as the network input, where the input feature size of single-phase electrical signals is $T \times T \times 1$, analogous to grayscale. The input feature size of three-phase electrical signals is $T \times T \times 3$, analogous to RGB color pictures. The network architecture uses fully connected layers to integrate high-dimensional features trained by Swin-T and then output categories. A Swin-T network with load identification capability is obtained through reverse propagation and multiple iterations.

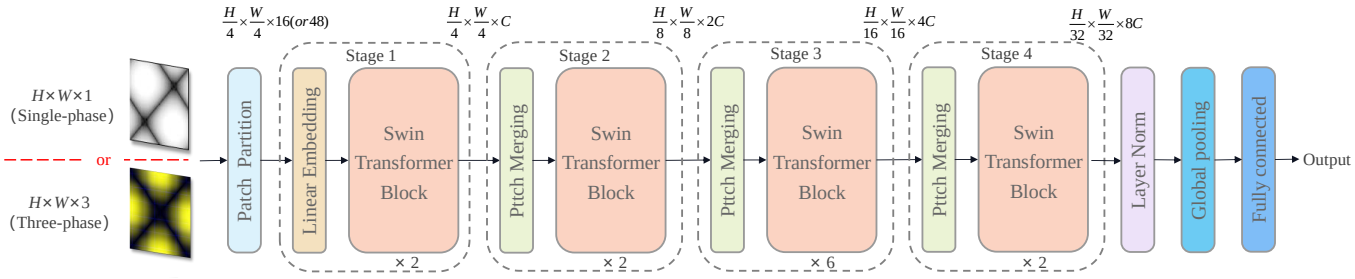


Figure 8. Main structure of Swin-T network for load identification.

3. Experimental Design

3.1. Datasets

The datasets LILACD (Laboratory-measured Industrial Load of Appliance Characteristics dataset) [33], PLAID (Plug Load Appliance Identification Dataset) [34], WHITED (Worldwide Household and Industry Transient Energy Data Set) [35], and HASED (Household Appliance Short-term Event Dataset) were applied to the proposed method. The LILACD dataset contains three-phase aggregated current and voltage measurements sampled at 50 KHz for 16 different loads covering industrial and domestic scenarios. To obtain more accurate measurements, most appliances were captured in different states such as heating temperature for hair dryers, different resistances for the resistor, or power usage for the motors. The PLAID contains single-phase aggregation voltages and currents for 12 different household appliances in 56 households in Pittsburgh, Pennsylvania, USA. This dataset is sampled at 30 KHz and contains 1478 measurements, obtained when multiple devices were active at the same time. Each appliance type is represented by dozens of different instances of varying make/models. For each appliance, three to six measurements were collected. These measurements were then postprocessed to extract a 1 s window containing both the steady-state operation and the start-up transient (when available). The WHITED consists of meter current and voltage measurements recorded at a sampling rate of 44.1 KHz in home and small industrial environments. It is based on using a sound card with an analog-to-digital converter to obtain data and record the electrical signal data 100 ms before and 5 s after the start-ups of each appliance.

To further verify the effectiveness of the method, we collected electrical signal data based on the self-developed equipment in our laboratory, as shown in Figure 9. We manually started-up seven types of electrical appliances and collected electrical signal data using the hardware shown. Then, the 1 ms electrical signal data before and after each appliance's start-up were extracted using the event detection algorithm proposed by [36]. To further verify the effectiveness of the method, we collected electrical signal data based on the self-developed equipment in our laboratory. The acquisition frequency of this device was 6.4 KHz, and the dataset HASED, which contains 7 types of loads and a total of 6938 instances, was obtained.

We removed the load category with too few samples in the above datasets and cleaned the samples with wrong labels and missing labels. Finally, the proportion of internal load samples in each dataset used for the experiment is shown in Table 1.

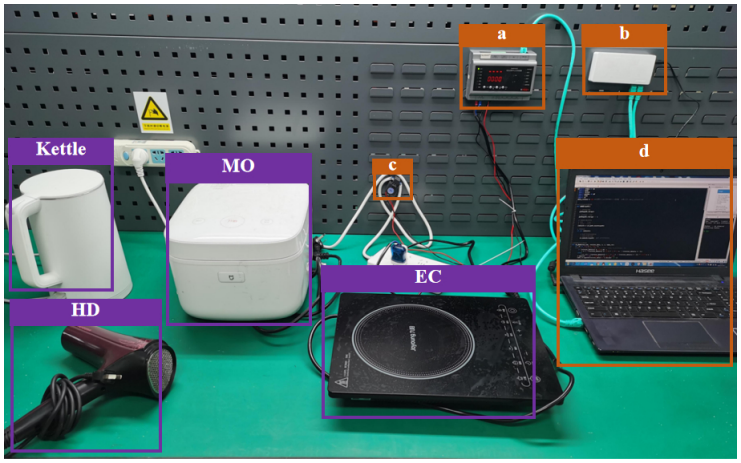


Figure 9. Acquisition environment and hardware connection of HASED dataset (a is used for data collection and transmission, b can realize network sharing, c is an electrical signal transformer, and d is a data receiving and processing terminal).

Table 1. Number of samples for each type of load within different datasets.

PLAID		LILACD		WHITED		HASED	
Load	**Amount**	**Load**	**Amount**	**Load**	**Amount**	**Load**	**Amount**
CFL	114	1-PAM	124	Charger	70	VC	916
Bulb	140	3-PAM	74	Fan	60	MO	1579
WK	128	Bulb	164	GC	40	Iron	1009
Fan	124	CM	74	HD	70	kettle	82
AC	114	Drilling	76	Iron	50	Fan	930
HD	122	Dumper	84	Kettle	89	EC	1192
LC	134	FL	80	Light	90	HD	1230
SI	162	FCS-3-2x	80	Bulb	70		
Fridge	104	HD	62	Mixer	40		
VC	128	Kettle	72	PS	40		
CM	116	RC	80	Toaster	40		
FD	92	RF	56	TV	40		
		Resistor	80	VC	40		
		S3A	84	WH	40		
		S3A-2x	80				
		VC	54				
total	1478	total	1324	total	779	total	6938

CFL (CompactFluorescentLamp), WK (WaterKettle), LC (LaptopCharger), FD (FridgeDefroster), 1-PAM (1-phase-async-motor), 3-PAM (3-phase-async-motor), CM (CoffeeMaker), FL (Fluorescentlamp), SI (SolderingIron), FCS-3-2x (Freq-conv-squirrel-3-2x), HD (HairDryer), RC (Raclette), RF (Refrigerator), S3A (Squirrel-3-async), S3A-2x (Squirrel-3-async-2x), VC (VacuumCleaner), GC (GameConsole), PS (PowerSupply), WH (WaterHeater), MO (MicrowaveOven), EC (ElectricCooker), AC (Airconditioner).

3.2. Method Comparison and Parameter Settings

To verify the effectiveness of the method, the recognition effects of ASRP, V-I trajectory, spectrogram, and CWT were compared. For three-phase electrical signal data, we converted each phase data to ASRP first and then merged them into a single load feature. Let $\varepsilon_k = \max(I_k) \times 10\%$ [37] to make rational use of the non-linear rendering effect of RP. Besides, complex Morlet wavelet was selected as the fundamental wave for CWT feature extraction [16]. In addition, please refer to [31] for the selection of various parameters of the identification network Swin-T, as shown in Table 2.

Table 2. Swin-T network hyper-parameter selection.

Learning Rate	Batch Size	Epoch	Optimizer
0.00001	4	100	adamW

The above network uses the standard open-source framework pytorch (https://pytorch.org (accessed on 12 March 2022)) of the Python platform (https://www.python.org (accessed on 12 March 2022)) for program development. The implemented hardware specifications are a 2.00 GHz AMD EPYC 7713 64-Core Processor CPU, an NVIDIA A100 PCIE 40 GB GPU, and 64.00 GB RAM.

K-fold Cross-Validation (K-CV) is utilized for training and generating to verify the control generation capability of the given model for unknown scenario patterns. K-CV is implemented by randomly dividing the original dataset into K groups. In our experiment, K was set to 10. Then, each subset of data was used as a validation set, while the remaining $K-1$ subsets of data were utilized as a training set, resulting in K models. The final score for the load identification method is the average classification score of the final validation set of these K models.

3.3. Performance Metrics

The macro-averaged F_1 score (F_{MA}), Kappa Index of Agreement (KIA), and Matthew's Correlation Coefficient (MCC) were employed to quantify the classification performance of the load identification approach. In addition, we built a confusion matrix that contrasted the network output to the actual classification that was known.

The F_1 score can thoroughly evaluate the classification efficiency. However, it is affected by the sample size of each category. Thus, F_{MA} is utilized to evaluate the classification effect in order to compensate for the evaluation mistake induced by the sample size.

$$F_{MA} = 100 \times \frac{1}{M} \sum_{i=1}^{M} F_1^i \tag{8}$$

M represents the total number of load categories and F_1^i represents the harmonic mean of the precision and recall of the i-th load. KIA is an index that can estimate the overall consistency and classification consistency. It can be used to evaluate the accuracy of the multi-classification model. The value of KIA is positively correlated with the classification effect of the model. KIA is defined as follows:

$$\mathrm{KIA} = \frac{p_0 - p_c}{1 - p_c} \tag{9}$$

$$\begin{aligned} p_0 &= \frac{\sum_{i=1}^{M} x_{ii}}{N} \\ p_c &= \frac{\sum_{i=1}^{M} x_{i+} x_{+i}}{N^2} \end{aligned} \tag{10}$$

where x_{ii} represents the elements on the diagonal of the confusion matrix, x_{i+} represents the sum of all elements in line i, x_{+i} represents the sum of all elements in column i, and n represents the total amount of samples.

p_0 is the proportion of observed accuracy or consistent units; p_c is the proportion of chance-consistent or expected chance-consistent units.

MCC provides a balanced performance measure of the quality of classification algorithms [38]. MCC is usually used to evaluate binary classification. For multi-classification tasks, given a confusion matrix C, MCC can be defined as follows:

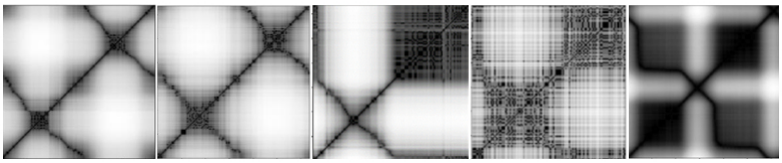
$$\mathrm{MCC} = \frac{c \times s - \sum_k^M p_k \times t_k}{\sqrt{\left(s^2 - \sum_k^M p_k^2\right) \times \left(s^2 - \sum_k^M t_k^2\right)}} \tag{11}$$

where $t_k = \sum_i^M C_{ik}$, $p_k = \sum_i^M C_{ki}$, $c = \sum_i^M C_{ii}$, $s = \sum_i^M \sum_j^M C_{ij}$. A score between −1 and +1 is returned from MCC. This score correlates positively with prediction performance, which represents perfect prediction when it is +1.

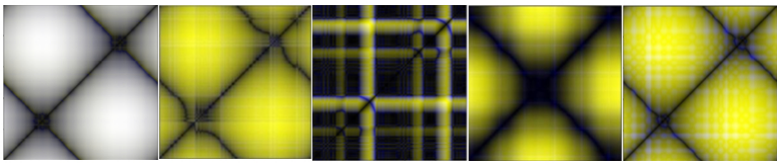
4. Result and Discussion

The ASRP features obtained after feature extraction of various loads are shown in Figure 10. Figure 11 shows that the electric signal curves of WK and CM loads are highly similar. If RP feature extraction is adopted, key feature details will be lost, which is unfavorable to load identification. The ASRP feature proposed in this paper can effectively improve the feature discrimination of similar loads by using adaptive scaling to preserve and amplify the feature differences.

The collected features of each dataset were fed into the created Swin-T network for training and testing. The final recognition results are displayed in Tables 3 and 4, and the respective feature effects of each evaluation index are compared in Figure 12.

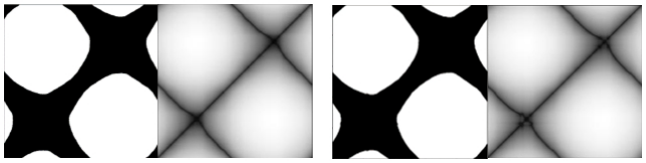


(**a**)



(**b**)

Figure 10. ASRP thermal maps of different loads obtained by feature extraction. (**a**) Single-phase load (VC, Mo, iron, EC, kettle). (**b**) Three-phase load (FL, Dumper, Drilling, CM, Resistor).



(**a**) (**b**)

Figure 11. Comparison of RP and ASRP feature display effects. (**a**) RP and ASRP of WK. (**b**) RP and ASRP of CM.

Table 3. Number of successful identifications of load events in each dataset (Event number is the total number of events for each type of load in each dataset. ASRP, RP, CWT, V-I trajectory, Spectrogram are the number of events accurately identified after the corresponding feature extraction method is used).

Dataset	Event Number	ASRP	RP	CWT	V-I Trajectory	Spectrogram
PLAID	1478	1453 (98.3%)	1436 (97.1%)	1163 (78.7%)	1232 (83.3%)	1090 (73.7%)
LILACD	1324	1295 (97.8%)	1211 (91.5%)	876 (66.2%)	854 (64.5%)	731 (55.2%)
WHITED	779	762 (97.8%)	760 (97.5%)	424 (54.4%)	467 (59.9%)	464 (59.6%)
HASED	6938	6885 (99.2%)	6881 (99.2%)	6717 (96.8%)	6809 (98.1%)	6566 (94.6%)

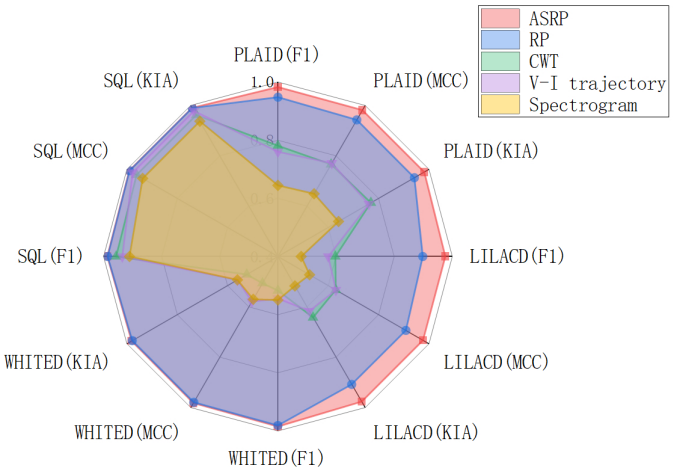


Figure 12. Comparison of the recognition results obtained by different load feature extraction methods for each dataset.

Table 4. Comparison of recognition scores obtained with different load feature extraction methods for each dataset.

Dataset	Method	F1	MCC	KIA
PLAID	ASRP	0.9826	0.9814	0.9815
	RP	0.9479	0.9421	0.9427
	CWT	0.7811	0.7658	0.7695
	V-I trajectory	0.7603	0.7613	0.7708
	spectrogram	0.6444	0.6408	0.6490
LILACD	ASRP	0.9760	0.9763	0.9764
	RP	0.8981	0.9077	0.9082
	CWT	0.5978	0.6299	0.6421
	V-I trajectory	0.5731	0.6166	0.6302
	spectrogram	0.4798	0.5169	0.5256
WHITED	ASRP	0.9846	0.9815	0.9817
	RP	0.9813	0.9788	0.9791
	CWT	0.5171	0.5061	0.5237
	V-I trajectory	0.5445	0.5605	0.5790
	spectrogram	0.5497	0.5609	0.5700
HASED	ASRP	0.9870	0.9904	0.9904
	RP	0.9858	0.9897	0.9895
	CWT	0.9567	0.9616	0.9620
	V-I trajectory	0.9356	0.9775	0.9776
	spectrogram	0.9108	0.9368	0.9374

The numerical results indicate that the performance of RP and ASRP is always superior to that of CWT, V-I trajectory, and spectrogram. The V-I trajectory and spectrogram provide little information and lack intricate contours and details. In addition, the location of their vital information is rather concentrated. Although CWT is mixed with time–frequency information, it cannot reflect obvious feature differences for loads with stable waveform. Due to the above factors, they are unsuitable for high-dimensional feature representation by the graph recognition network. In contrast, RP and ASRP features reflect the non-linear characteristics of time-series signals. Moreover, their feature information is distributed uniformly with various contours and dense pixel details. This can successfully increase the ability of feature discrimination between similar loads and reduce the impact of data volume on the recognition effect. For the WHITED, LILACD, and PLAID datasets in the experiment, the recognition scores of V-I trajectory and spectrogram are all less than 0.80, while the recognition scores of RP and ASRP are both 0.90 and higher. As the amount of

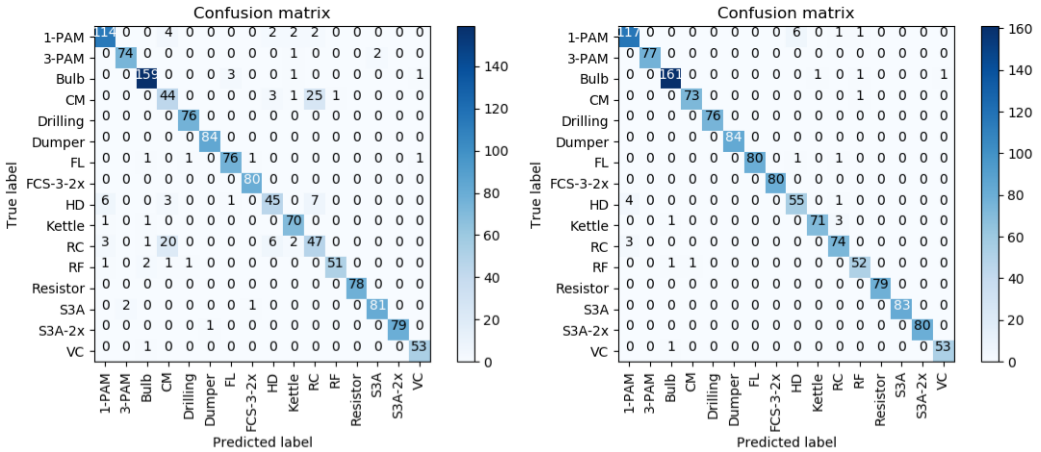
data—such as in the HASED dataset—increases, the performances of CWT, V-I trajectory, and spectrogram improve. Nevertheless, the recognition effects of RP and ASRP continue to offer advantages.

RP and ASRP features can effectively improve the feature discrimination of different loads, so as to reduce the dependence of deep learning on the amount of data and the time cost caused by data acquisition in the actual deployment of NILM system. In addition, it can also enhance the analytical capability of NILM by further exploiting the advantages of deep learning.

ASRP has more information compared with the binary numerical characteristic RP. In addition, it eliminates the tedious hyper-parameter formulation stages and improves the discrimination of similar loads by using the electrical signal differences of various loads to perform adaptive scaling. The ASRP scores exceeded 0.97 in each of the four datasets used in the studies. Compared with RP, the recognition scores for the PLAID dataset increased from 0.94 to 0.98. Moreover, using the less focused three-phase dataset LILACD, the confusion matrix (Figure 13) indicates that it is difficult to distinguish RC and CM loads with similar electrical signals, even with RP characteristics. However, ASRP can substantially improve the feature discrimination of the approximation load, increasing the score for recognition from 0.90 to 0.97.

The results show that ASRP can quickly adapt to heavy load scenarios with similar loads. In addition, ASRP can achieve rapid configuration and maintenance of the algorithm by removing parameter adjustment to effectively improve the stability of the NILM system compared with other methods. Furthermore, regardless of using single-phase or three-phase electrical signals, ASRP can obtain excellent recognition effect, indicating that ASRP can be applied to all kinds of home/industrial environments so as to actively promote the development of home/industrial information Internet of Things.

In summary, traditional V-I trajectory and spectrogram features are no longer adequate for classifying equipment in residential and industrial applications. In contrast, ASRP effectively increases the differentiation between load characteristics and the uniqueness of each load. It is also widely used for single-phase and three-phase load detection.



(**a**) (**b**)

Figure 13. Showing the load identification results obtained by different feature extraction methods for LILACD dataset through confusion matrix. (**a**) Confusion Matrix (LILACD-RP). (**b**) Confusion Matrix (LILACD-ARM).

5. Conclusions

This paper presented a promising NILM method based on deep learning that can adaptively characterize load features. To enhance the feature difference between similar loads, an adaptive scaling of the threshold-free RP is utilized for feature extraction regardless of formulating hyper-parameters. Moreover, load recognition can be achieved through multi-

head attention and the shifted window mechanism built-in to Swin-Transformer efficiently and accurately. The proposed solution has been widely validated on four real, measured load signal datasets, including industrial and domestic power situations, covering single-phase and three-phase electrical signals. Compared with V-I trajectory, spectrogram, and RP, the proposed NILM method based on adaptive RP has a higher accuracy according to the numerical results.

In practical application, the adaptive parameter adjustment feature can provide stronger adaptability for NILM technology and reduce the maintenance cost in later periods. Furthermore, combining ASRP with deep learning can effectively reduce the demand for sample data, thus reducing data collection and time costs. In addition, it adapts to single-phase and three-phase electrical signals, enabling NILM to be applied to various power consumption scenarios.

Author Contributions: Conceptualization, methodology, Y.S. and X.Z.; software, validation, Y.S.; data curation, manuscript review, and revision, F.Z. and Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Key Projects of Science and Technology Plan of Zhejiang Province under Grant 2021C01144.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Angelis, G.F.; Timplalexis, C.; Krinidis, S.; Ioannidis, D.; Tzovaras, D. NILM applications: Literature review of learning approaches, recent developments and challenges. *Energy Build.* **2022**, *261*, 111951. [CrossRef]
2. Batra, N. Systems and Analytical Techniques towards Practical Energy Breakdowns for Homes. Ph.D. Thesis, Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi), Delhi, India, 2017.
3. He, K.; Stankovic, L.; Liao, J.; Stankovic, V. Non-Intrusive Load Disaggregation Using Graph Signal Processing. *IEEE Trans. Smart Grid* **2018**, *9*, 1739–1747. [CrossRef]
4. Carrie Armel, K.; Gupta, A.; Shrimali, G.; Albert, A. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* **2013**, *52*, 213–234. [CrossRef]
5. Athanasiadis, C.; Doukas, D.; Papadopoulos, T.; Chrysopoulos, A. A Scalable Real-Time Non-Intrusive Load Monitoring System for the Estimation of Household Appliance Power Consumption. *Energies* **2021**, *14*, 767. [CrossRef]
6. Gopinath, R.; Kumar, M.; Prakash Chandra Joshua, C.; Srinivas, K. Energy management using non-intrusive load monitoring techniques – State-of-the-art and future research directions. *Sustain. Cities Soc.* **2020**, *62*, 102411. [CrossRef]
7. Lin, J.; Luan, W.; Liu, B. A Novel Non-intrusive Arc Fault Detection Method for Low-Voltage Customers. In Proceedings of the 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), Chongqing, China, 8–11 April 2021; pp. 84–88. [CrossRef]
8. Hart, G. Nonintrusive appliance load monitoring. *Proc. IEEE* **1992**, *80*, 1870–1891. [CrossRef]
9. de Aguiar, E.L.; Lazzaretti, A.E.; Mulinari, B.M.; Pipa, D.R. Scattering Transform for Classification in Non-Intrusive Load Monitoring. *Energies* **2021**, *14*, 6796. [CrossRef]
10. Mariscotti, A. Non-Intrusive Load Monitoring Applied to AC Railways. *Energies* **2022**, *15*, 4141. [CrossRef]
11. Jia, D.; Li, Y.; Du, Z.; Xu, J.; Yin, B. Non-Intrusive Load Identification Using Reconstructed Voltage–Current Images. *IEEE Access* **2021**, *9*, 77349–77358. [CrossRef]
12. Wang, S.; Chen, H.; Guo, L.; Xu, D. Non-intrusive load identification based on the improved voltage-current trajectory with discrete color encoding background and deep-forest classifier. *Energy Build.* **2021**, *244*, 111043. [CrossRef]
13. Yang, C.C.; Soh, C.S.; Yap, V.V. A systematic approach in load disaggregation utilizing a multi-stage classification algorithm for consumer electrical appliances classification. *Front. Energy* **2019**, *13*, 386–398. [CrossRef]
14. Himeur, Y.; Alsalemi, A.; Bensaali, F.; Amira, A. Effective non-intrusive load monitoring of buildings based on a novel multi-descriptor fusion with dimensionality reduction. *Appl. Energy* **2020**, *279*, 115872. [CrossRef]
15. Chaurasiya, H. Time-Frequency Representations: Spectrogram, Cochleogram and Correlogram. *Procedia Comput. Sci.* **2020**, *167*, 1901–1910. [CrossRef]
16. Duarte, C.; Delmar, P.; Goossen, K.W.; Barner, K.; Gomez-Luna, E. Non-intrusive load monitoring based on switching voltage transients and wavelet transforms. In Proceedings of the 2012 Future of Instrumentation International Workshop (FIIW) Proceedings, Gatlinburg, TN, USA, 8–9 October 2012; pp. 1–4. [CrossRef]

17. Wenninger, M.; Bayerl, S.P.; Maier, A.; Schmidt, J. Recurrence Plot Spacial Pyramid Pooling Network for Appliance Identification in Non-Intrusive Load Monitoring. In Proceedings of the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA, 13–16 December 2021; pp. 108–115. [CrossRef]
18. Faustine, A.; Pereira, L. Improved Appliance Classification in Non-Intrusive Load Monitoring Using Weighted Recurrence Graph and Convolutional Neural Networks. *Energies* **2020**, *13*, 3374. [CrossRef]
19. Bouhouras, A.S.; Gkaidatzis, P.A.; Panagiotou, E.; Poulakis, N.; Christoforidis, G.C. A NILM algorithm with enhanced disaggregation scheme under harmonic current vectors. *Energy Build.* **2019**, *183*, 392–407. [CrossRef]
20. Yan, L.; Sheikholeslami, M.; Gong, W.; Tian, W.; Li, Z. Challenges for real-world applications of nonintrusive load monitoring and opportunities for machine learning approaches. *Electr. J.* **2022**, *35*, 107136. [CrossRef]
21. Berrettoni, G.; Bourelly, C.; Capriglione, D.; Ferrigno, L.; Miele, G. Preliminary Sensitivity Analysis of Combinatorial Optimization (CO) for NILM Applications: Effect of the Meter Accuracy. In Proceedings of the 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), Naples, Italy, 6–9 September 2021; pp. 486–490. [CrossRef]
22. Guo, Y.; Xiong, X.; Fu, Q.; Xu, L.; Jing, S. Research on non-intrusive load disaggregation method based on multi-model combination. *Electr. Power Syst. Res.* **2021**, *200*, 107472. [CrossRef]
23. Gurbuz, F.B.; Bayindir, R.; Vadi, S. Comprehensive Non-Intrusive Load Monitoring Process: Device Event Detection, Device Feature Extraction and Device Identification Using KNN, Random Forest and Decision Tree. In Proceedings of the 2021 10th International Conference on Renewable Energy Research and Application (ICRERA), Istanbul, Turkey, 26–29 September 2021; pp. 447–452. [CrossRef]
24. Li, Y.; Yang, Y.; Sima, K.; Li, B.; Sun, T.; Li, X. Non-intrusive load monitoring based on harmonic characteristics. *Procedia Comput. Sci.* **2021**, *183*, 776–782. [CrossRef]
25. Davies, P.; Dennis, J.; Hansom, J.; Martin, W.; Stankevicius, A.; Ward, L. Deep Neural Networks for Appliance Transient Classification. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8320–8324. [CrossRef]
26. Wu, Q.; Wang, F. Concatenate Convolutional Neural Networks for Non-Intrusive Load Monitoring across Complex Background. *Energies* **2019**, *12*, 1572. [CrossRef]
27. He, X.; Dong, H.; Yang, W.; Hong, J. A Novel Denoising Auto-Encoder-Based Approach for Non-Intrusive Residential Load Monitoring. *Energies* **2022**, *15*, 2290. [CrossRef]
28. Song, J.; Wang, H.; Du, M.; Peng, L.; Zhang, S.; Xu, G. Non-Intrusive Load Identification Method Based on Improved Long Short Term Memory Network. *Energies* **2021**, *14*, 684. [CrossRef]
29. Zheng, Z.; Chen, H.; Luo, X. A Supervised Event-Based Non-Intrusive Load Monitoring for Non-Linear Appliances. *Sustainability* **2018**, *10*, 1001. [CrossRef]
30. Eckmann, J.P.; Kamphorst, S.O.; Ruelle, D. Recurrence Plots of Dynamical Systems. *Europhys. Lett.* **1987**, *4*, 973–977. [CrossRef]
31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [CrossRef]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
33. Kahl, M.; Krause, V.; Hackenberg, R.; Haq, A.U.; Horn, A.; Jacobsen, H.A.; Kriechbaumer, T.; Petzenhauser, M.; Shamonin, M.; Udalzow, A. Measurement system and dataset for in-depth analysis of appliance energy consumption in industrial environment. *Tech. Mess.* **2019**, *86*, 1–13. [CrossRef]
34. Medico, R.; De Baets, L.; Gao, J.; Giri, S.; Kara, E.; Dhaene, T.; Develder, C.; Bergés, M.; Deschrijver, D. A voltage and current measurement dataset for plug load appliance identification in households. *Sci. Data* **2020**, *7*, 49. [CrossRef]
35. Kahl, M.; Haq, A.U.; Kriechbaumer, T.; Jacobsen, H.A. Whited-a worldwide household and industry transient energy data set. In Proceedings of the 3rd International Workshop on Non-Intrusive Load Monitoring, Vancouver, BC, Canada, 14–15 May 2016; pp. 1–4.
36. Zhang, F.; Qu, L.; Dong, W.; Zou, H.; Guo, Q.; Kong, Y. A Novel NILM Event Detection Algorithm Based on Different Frequency Scales. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [CrossRef]
37. Matassini, L.; Kantz, H.; Hołyst, J.; Hegger, R. Optimizing of recurrence plots for noise reduction. *Phys. Rev. E* **2002**, *65*, 021102. [CrossRef]
38. Pereira, L.; Nunes, N. A comparison of performance metrics for event classification in Non-Intrusive Load Monitoring. In Proceedings of the 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm), Dresden, Germany, 23–27 October 2017; pp. 159–164. [CrossRef]