This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Non-Intrusive Load Identification Method based on Self-Supervised Regularization

Ruifeng Zhao[1], Jiangang Lu[1], Bo Liu[2], Zhiwen Yu[1], Yanru Ren[2] and Wenjie Zheng[1].

[1]Electric Power Dispatching and Control Center, Guangdong Power Grid, Guangzhou 510600, China
[2]School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Corresponding author: Ruifeng Zhao(ruifzhao@126.com), Bo Liu (liubo@tju.edu.cn).

This work was funded by the National Key Research and Development Program of China (No.2022YFB2403800) and the Science and Technology Project of State Grid Corporation Headquarters(No.5400-202255274A-2-0-XG).

ABSTRACT Non-intrusive load monitoring is a novel and cost-effective technology for monitoring details of electricity consumption and identifying the operating status of appliances. It supports the construction of the energy internet and big data on electricity consumption in smart cities. However, one of the most challenging problems in this area is that machine learning algorithms often require large amounts of labeled data. In this paper, a non-intrusive load monitoring model based on the Self-supervised Regularization is proposed. The model reduces the pre-processing stage compared to the traditional methods. We make full use of the unlabeled data by using them to generate proxy labels to participate in the model training together with the true labels. We performed experiments on the common data set PLAID to compare performance with the existing method Mean Teacher and CoMatch. The experimental results show that： (1) when using all labeled data, the model with self-supervised regularization significantly improves the traditional supervised classifier with a recognition accuracy of 0.965. (2) when coupled with unlabeled data, our model produces good semi-supervised performance. It is highly competitive with current state-of-the-art Mean Teacher and Contrastive Learning.

INDEX TERMS non-intrusive load monitoring, semi-supervised learning, self-supervised regularization, proxy label

I. INTRODUCTION

Energy is the cornerstone of human survival and development. How to save energy and improve energy efficiency has become the focus of attention. As one of the main forms of energy consumption, electricity is closely related to people's production and life. In recent years, with the reform of electricity market, improving the efficiency of electricity utilization and reducing the waste of electricity resources have become the priority. The proposal of load monitoring technology enables the monitoring of all types of appliances in residential commercial, and industrial buildings. Load monitoring technology is divided into intrusive load monitoring (ILM) and non-intrusive load monitoring (NILM). ILM can provide fast and accurate information about the customer's appliance. However, this results in a range of shortcomings such as time consuming, complex wiring, high cost and so on. In the 1980s, non-intrusive load monitoring [1] was proposed which differs from traditional ILM in that it only needs to install sensors in the power bus for monitoring, through the bus terminal voltage, current, power, and other parameters, to achieve the purpose of monitoring the state of each appliance and energy loss. Because this method only needs to install sensors at the power bus to collect voltage and current information, or directly integrated into the smart meter, it has the advantages of easy maintenance and low cost. In recent years, non-intrusive load monitoring has become a research hotspot.

Non-intrusive Load Monitoring (NILM) is critical in the context of energy efficiency, energy security, and power consumption. NILM allows users to identify energy consumption of domestic, commercial, or industrial devices. By identifying high energy consuming device or patterns of energy consumption, users can take steps to reduce energy waste and improve energy efficiency. In addition, NILM can detect unusual energy consumption patterns, such as unusual equipment behavior. By identifying these issues in a timely manner, the safety of the energy system can be enhanced and the risk of electrical accidents can be reduced. NILM can also be used to help users optimize power consumption, for example by using power equipment during peak and valley times. This can help balance the load on the power grid and reduce peaks in power demand, thereby reducing stress on the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

power system and improving grid stability. Studies have shown that 9% ~ 20% of electricity can be saved through the appliance-level power consumption information [2].

In the monitoring process, the NILM device needs to continuously acquire the user's aggregated data and extract load signatures from it and then realize the load decomposition according to the load signature library, using the mapping relationship between the load signature and the operating state of the corresponding appliance. The two main target of load decomposition are the energy disaggregation and appliance identification. The problem this paper addressed is the appliance identification problem. The extracted features are compared with the load signature template in the existing load signature library, and when they reach a certain degree of similarity, the corresponding appliances are identified. How to improve the identification accuracy is now the main concern.

Load signatures are the main basis for load identification, which are divided into steady-state signatures and transient signatures [3-7]. Steady-state signatures are simpler to measure and process, and easy to extract. However, they are limited when recognizing resistive loads with similar power. In comparison, transient signatures have less information overlap and higher recognition. However, transient features have higher requirements on sampling frequency and sampling devices. For state identification of residential appliances, we usually use steady-state signatures to fulfill the requirements.

V-I trajectory can provide more comprehensive information about the state of appliance than other steady-state signatures. Since the V-I trajectory uses continuous current and voltage sampling data, combined with time and amplitude information, it more accurately reflects the state of the appliance. For similar appliances, V-I trajectory is more distinguishable than other steady-state features, which can identify appliances at the higher spectral bandwidths and the richer information perspectives. The method of Self-supervised Regularization proposed in this paper is based on V-I trajectory.

In the field of NILM, we need to obtain the user's electricity consumption information as the training data for the model, however, we cannot obtain this information directly from the Internet. Generally, we need to install sensors on appliances inside user for real-time monitoring, and obviously it is difficult and time-consuming to obtain enough appliance-level training data. In addition, we need to annotate the acquired data, but there is no good technology to realize automatic annotation of data, which requires a lot of manual and cost. Therefore, it is necessary to use small-size datasets with annotations for learning in the field of NILM.

In recent years, machine learning (ML) models have been widely used in modeling, designing, and predicting energy systems and have achieved good results in the field of solving classification problems. In ref [8], three machine learning algorithms: Deep Neural Network, Convolutional Neural Network, and Subtask Gated Network were applied to different datasets, which verified the dependence of machine learning algorithms on the dataset size. In ref [9], a deep Q-learning framework to ensure secure operation of modern power-grid systems is proposed. Ref [10] presents the state of the art of ML models used in energy systems. It can be observed that NILM algorithms are classified as supervised learning, unsupervised learning, and semi-supervised learning according to whether labeled data is required. The supervised learning method requires the use of labeled data for model training, which usually uses classifiers or neural networks for load identification. The advantage is that the identification accuracy is high, but a large amount of data needs to be manually labelled, and the model needs to be retrained for new appliance types. The unsupervised learning method does not need to use labeled data for model training, and realizes load identification by clustering or dimensionality reduction of data and analyzing data features. The advantage is that there is no need to label data samples, but the accuracy may be lower, and different states of the same appliance may be misjudged as different appliances. Load identification algorithm based on semi-supervised learning combining supervised learning and unsupervised learning, utilizes unlabeled data for model optimization while using a small amount of labeled data for model training, which has a relatively higher accuracy but does not require many labeled data. In this paper, a new semi-supervised method is proposed to achieve the use of unlabeled data.

In semi-supervised field, contrastive learning pretext-learning has demonstrated good performance. Ref [11] describes SimCLR: a simple framework for contrast learning of visual representations. It simplifies recently proposed contrastive self-supervised learning algorithms. Ref [12] proposed a Graph Comparative Learning (GraphCL) framework with four graph augmentation types designed, for learning unsupervised representations of graph data. Ref [13] proposes CoMatch, a new semi-supervised learning method that unifies mainstream comparison methods and addresses their limitations. CoMatch jointly learns two representations of the training data, class probabilities and low-dimensional embeddings. In this paper, the proposed method is compared with the CoMatch method.

To summarize the contribution of this paper, there are three main points:

- For NILM, a new Self-supervised Regularization (SESR) method is proposed in this paper, which is a combination of consistent regularization method and traditional self-supervised method.
- The SESR method generates proxy labels for the unlabeled data, and the self-supervised loss terms generated by the proxy labels are used to correct the model together with the supervised losses generated by the true labels.
- In this paper, experimental validation on the PLAID dataset is carried out to produce better semi-supervised performance when compared with the Mean Teacher method. In the context of image classification, it can make use of rich

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

unlabeled data to significantly improve the classifier in a limited set of labeled data.

The rest of the paper is presented as follows: the V-I trajectory signature acquisition process and the proposed Self-supervised Regularization method are presented in Section II. The experimental results and analysis are presented in Section III. The conclusion of this paper is in Section IV.

II. THE PROPOSED APPROACH

In this section, the process of acquiring V-I trajectory signature and encoding V-I trajectories into color images is first described, followed by existing semi-supervised methods, and finally leading to the method proposed in this paper: a non-intrusive load identification model based on Self-supervised Regularization.

A. V-I Trajectory Signature Acquisition

As a bridge between computer vision and NILM, the V-I trajectory features have been widely used, so in this paper, we use V-I trajectory features of appliances as input to the model.

The V-I trajectory is a set of parameters obtained by plotting and processing V-I signatures from sampled data of currents and voltages in a circuit. V-I trajectory has the potential to be an intermediate field between computer vision and NILM. The V-I trajectory of each appliance can be plotted from its normalized current and voltage. Once the V-I signatures have been extracted, different algorithms can be used to identify the appliance.

We refer to the method in the reference [14] to extract the V-I trajectory:

- Extract the voltage and current waveforms for T cycles before and after the event (A switch in the operating states of an appliance is referred to as an event).
- Linear interpolation of voltage waveform and Elmi interpolation of current waveform.
- Define the voltage and current values of the appliance：

$$\boldsymbol{v} = \frac{\boldsymbol{v}_{on} + \boldsymbol{v}_{off}}{2} \tag{1}$$

$$\boldsymbol{i} = \boldsymbol{i}_{off} - \boldsymbol{i}_{on} \tag{2}$$

Where, $\boldsymbol{v}_{on}$ and $\boldsymbol{i}_{on}$ are the voltage vectors and current vectors for T cycles before the event and consist of J observations. $\boldsymbol{v}_{off}$ and $\boldsymbol{i}_{off}$ are the voltage vectors and current vectors for T cycles after the event and consist of J observations. $\boldsymbol{v}$ and $\boldsymbol{i}$ respectively represent the voltae and current vectors composed of J sampling points in steady-stable period.

- Encode the V-I trajectory as a color image with HSV (Hue, Saturation, Value) color space [15].

The Hue is used to indicate the direction of the trajectory, H_j is used to represent the hue dimension encoding of the *jth* point in the trajectory as shown in the formula (3)：

$$H_j = \arg\left(\frac{v_{j+1} - v_j}{v_{\max}}, \frac{i_{j+1} - i_j}{i_{\max}}\right),\ j = 1, 2, 3 \ldots J-1 \tag{3}$$

Where, the function $\arg(\cdot)$ is a four-quadrant anti tangent to calculate the phase angle of two consecutive points in the trajectory. $v_{\max}$ and $i_{\max}$ are the maximum value of vector $\boldsymbol{v}$ and $\boldsymbol{i}$.

The Saturation is used to express the power factor of multiple cycles, as shown in the formula (4):

$$S = \frac{P_{active}}{P_{apparent}} = \frac{\frac{1}{J}\sum_{n=1}^{J} v_n \cdot i_n}{v_{rms} \cdot i_{rms}} \tag{4}$$

Where, J is the total number of sampling points. P_{active} and $P_{apparent}$ are active power and apparent power in one stable period obtained by $\boldsymbol{v}$ and $\boldsymbol{i}$, respectively. v_{rms} and i_{rms} are the effective value of load voltage $\boldsymbol{v}$ and current $\boldsymbol{i}$, respectively.

The Value is used to express the repeatability of the V-I trajectory, as shown in the formula (5):

$$V = \frac{1}{T}\sum_{t=1}^{T} W_t \tag{5}$$

Where, T is the number of cycles, W_t is the $2N \times 2N$ matrix generated by the binary trajectory mapping of voltage and current waveforms for T cycles [16].

To create a color image that humans can perceive, the matrices H, S and V can be thought of as equivalent red-green-blue colors, as shown in Fig. 1.

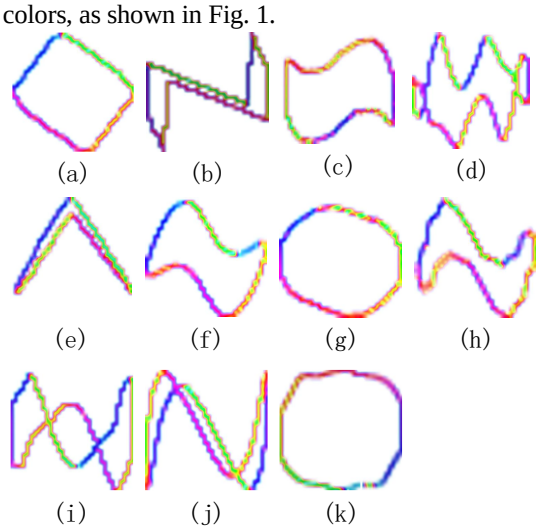


FIGURE 1. The color images generated on the PLAID dataset of (a) air conditioner, (b) compact fluorescent lamp, (c) fan, (d) fridge, (e) hairdryer, (f) heater, (g) incandescent light bulb, (h) laptop, (i) microwave, (j) vacuum and (k) washing machine.

B. Semi-supervised learning for NILM

Many methods have been proposed to learn representations from large amounts of unlabeled data to augment the limited true labels for the semi-supervised learning. We summarize two specific classes of methods relevant to the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

work in this paper: consistency regularization and self-supervised learning.

Consistency regularization performs model parameter tuning via a loss function. This method typically utilizes two parts of the data, labeled data and unlabeled data. Typically, the unlabeled data produces a loss that becomes the auxiliary, unsupervised consistency loss, which is formulated as a regularizer that will jointly adjust the model parameters with the supervised loss. Regularization methods are used to effectively utilize unlabeled data to reduce overfitting in semi-supervised learning. In short, classification models need to have the ability to provide consistent outputs for similar data points. Mean Teacher is a typical model based on consistency regularization.

Mean Teacher [28] is a relatively new semi-supervised learning algorithm that has recently emerged in the field of deep learning. It consists of two parts: teacher network and student network. The two networks have the same structure, only parameters are different. Its loss function is divided into two parts: supervised consistency loss and unsupervised cross-entropy loss. The supervised loss is used to ensure the performance of the model on labeled data, while the unsupervised loss ensures that the results of teacher network and student network are as similar as possible. Therefore, the teacher model is also considered as a regulariser to improve the performance of the student network.

Self-supervised learning is divided into two stages: (1) Pretrain self-supervised models with unlabeled data. This phase is like the unsupervised approach in that the goal is to learn rich visual representations from large-scale unlabeled images. (2) Fine-tuning the model with labeled data. In this stage, a small amount of labeled data is used to adjust the model obtained from the first stage to enhance the accuracy of the model. The simplest self-supervised task is to minimize the reconstruction error in a self-encoder to learn a low-dimensional feature representation where the proxy labels are the values of image pixels [17]. More complex self-supervised tasks, such as image restoration [18], grey-scale image color [19] and predictive image rotation [20], show impressive results in unsupervised visual feature learning.

C. The Proposed Method

The essence of Self-supervised Regularization (SESR) method is to generate proxy labels using unlabeled data, and the generated proxy labels are used as targets along with the labeled data to train the model. The overall objective function consists of a combination of supervised loss and self-supervised loss which is also known to us as auxiliary loss and reflects the contribution of unlabeled data. Compared to methods based on consistent regularization. It does not require additional hyperparameters to tune performance beyond the standard set used to train the neural network, which greatly simplifying the work process.

The block diagram of our proposed method Self-Supervised Regularization is shown in Fig. 2, where D_L denotes the labeled dataset and D_U denotes the unlabeled dataset. The goal of SESR is to train a prediction function $f_\theta(x)$ parameterized by θ, which uses a combination of D_L and D_U to obtain better prediction performance than using D_L alone.

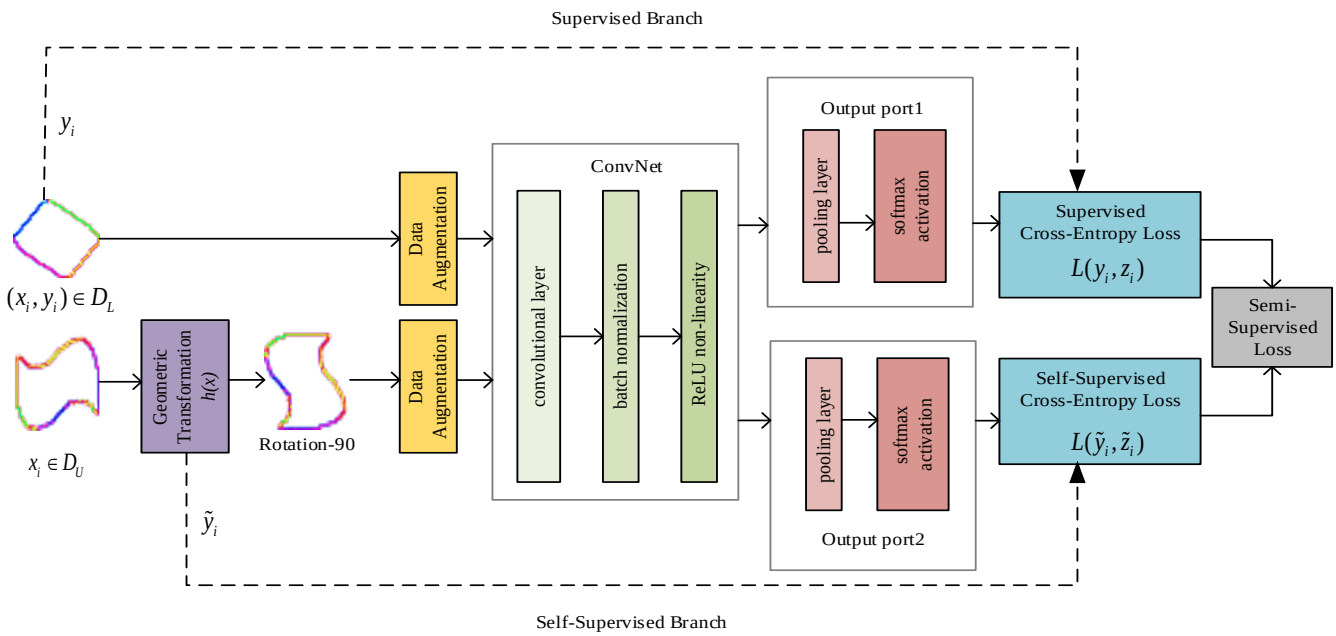


FIGURE 2. Self-supervised Regularization block diagram.

It can be seen that the whole model is divided into two branches: a supervised branch and a self-supervised branch. For supervised branch, the input is composed of target pairs. For the self-supervised branch, the input is the unlabeled data x_i, which generates six proxy labels through the function $h(x)$. An earlier study by Gidaris et al [23] defined the agent

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

labels as belonging to the rotation of images in the set of {0, 90, 180, 270} degrees, resulting in a four-way classification task. Their model performs well in the rotation identification task by learning salient visual features described in the image, such as the position, type and pose of the object. In this work, we extend the geometric transformation to include horizontal (left-right) and vertical (up-down) flipping.

As you can see in the model, we first sample the labeled and unlabeled data to ensure that the sampled two small batches have the same amount of data. Because there are many more unlabeled examples than labeled ones, the labeled examples will be used repeatedly in small batch processing. Forward propagation, where the two branches share a single prediction function $f_\theta(x)$ but have separate output heads, consists of a global average pooling layer and softmax activation. We chose a high-performance CNN framework: 13-layer max-pooling ConvNet [21-25] as prediction function $f_\theta(x)$. We faithfully follow the original specification of the ConvNet architecture, all of which have convolution layers followed by batch normalization [26] and ReLU nonlinearity [27]. The ConvNet architecture uses leaked ReLU, α=0.1. And it has approximately 3.13 million parameters.

For supervised branch, after propagation of $f_\theta(x)$, softmax prediction vector z_i is obtained. We use the true label y_i to calculate the supervised cross-entropy loss $L_{super}(y_i, z_i)$.

For self-supervised branch, after propagation of $f_\theta(x)$, a softmax prediction vector $\tilde{z}_i$ is obtained and we use the proxy labels $\tilde{y}_i$ generated from image rotation and flip to calculate the self-supervised cross-entropy loss $L_{super}(y_i, z_i)$. The objective function of the SESR model is defined as a weighted sum of the supervised loss and self-supervised cross-entropy loss:

$$L_{SESR} = L_{super}(y_i, z_i) + \omega L_{self}(\tilde{y}_i, \tilde{z}_i) \tag{6}$$

Where, ω is the regularization hyperparameter, and represents the relative contribution of self-supervision in the overall objective function. In the previous semi-supervised learning methods based on consistency regularization, such as the Mean Teacher, ω was defined as a consistency coefficient and heavily tuned on each dataset to obtain optimal performance. In this paper, it was verified after several experiments that ω=1 produces consistent results in all datasets and CNN architectures, so ω was set to 1. This also demonstrates that supervised and self-supervised losses are relatively balanced and compatible for image classification. The parameter θ is learned by back propagation.

III. EXPERIMENTS AND RESULTS

In this section, we first present the data preparation and evaluation metrics. Next, experiments using different proportions of unlabeled data on the dataset are conducted and compared with existing methods. Our experiments were implemented on Pytorch 1.8 and python 3.6, Intel(R) Xeon(R) W-2133 CPU, 3.60 GHz, 64.00 GB RAM, NVIDIA Quadro P2000. Due to the effects of toggling and unfiltered noise in the power system, we set the parameter T for the V-I trajectory extraction process to 10. The V-I trajectory size is set to 32*32.

A. Data preparation and Metrics

To validate the model, we used the Public Lighting Analysis and Identification Dataset (PLAID). Dataset PLAID is a public dataset for power monitoring and energy consumption analysis of embedded systems. The dataset, released in 2018 by researchers at Penn State, contains real-life electricity consumption data from 120 households, covering a two-year period from 2014 to 2015.

We chose the two commonly used F_{score} and F_{macro} as evaluation indicators to verify the feasibility of our proposed model. F_{score} is an indicator used to evaluate the performance of classification algorithms. It is the harmonic average of precision and recall. Therefore, it takes these two indicators into account to balance the performance of the model in different cases. As F_{score} approaches to one, the proposed method is considered to be an excellent non-intrusive load identification model. In contrast, model performance is considered unsatisfactory as F_{score} approaches to zero.

Precision is another common performance metric for non-intrusive load identification, which measures the proportion of *TP* among all predictions that are positive. We denote precision by *P*:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{7}$$

Where, N_{TP} indicates the number of samples for which the true value is positive and the model prediction is also positive, and N_{FP} indicates the number of samples for which the true value is negative and the model prediction is positive.

Recall refers to the proportion of *TP* samples among true labels to be positive. It is used to assess the detection ability of *TP* samples. We use *R* to represent the recall rate:

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{8}$$

Where, N_{FN} indicates the number of samples for which the true value is positive and the model predicted value is negative.

The two metrics, precision and recall, are clearly monotonic, they both increase as the performance of the classifier improves. In some cases, however, optimizing only one of these indicators may lead to deterioration of the other, so they need to be considered simultaneously:

$$F_{score} = 2 \cdot \frac{P \cdot R}{P + R} \tag{9}$$

Then, the macro mean F_{macro} was obtained by taking the average of all F_{score}

$$F_{macro} = \frac{1}{A} \sum_{i=1}^{A} F_{score} \tag{10}$$

Where, *A* is the total number of different appliance types.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

B. The supervised learning results (using all labels)

We followed a standard evaluation protocol for semi-supervised learning, in which we randomly sampled different parts of the training data as labeled examples while processing the entire training set, discarding all labeled information as the source of unlabeled data. We first analyzed the SESR model trained using only the labeled data, i.e. using only supervised analysis and discarding self-supervised analysis during training, and the results are shown in Fig. 3 and Fig. 4 We chose the confusion matrix for the analysis.

For all the analyses that follow, we use AC, CFL, Fan, FD, HD, HT, ILB, LT, MW, VAC and WM to represent the eleven appliances: air conditioner, compact fluorescent lamp, fan, fridge. hairdryer, heater, incandescent light bulb, laptop, microwave, vacuum cleaner, and washing machine.

The confusion matrix can reflect which appliances are confused with each other in the experiment. The horizontal coordinates in the figure indicate the predicted values of the model and the vertical coordinates indicate the true values of the model inputs. In the confusion matrix, the diagonal elements indicate the proportion of correct predictions, while the elements elsewhere indicate the proportion of incorrect predictions.

In the experiment, we choose the Mean Teacher method as the comparison term. A comparison of Mean Teacher with other non-intrusive load identification methods is given in the reference [28]. Mean Teacher is in a high position among the existing semi-supervised methods and it has an advantage in terms of accuracy compared to other semi-supervised methods, so the Mean Teacher method is chosen as the comparison method in this paper.

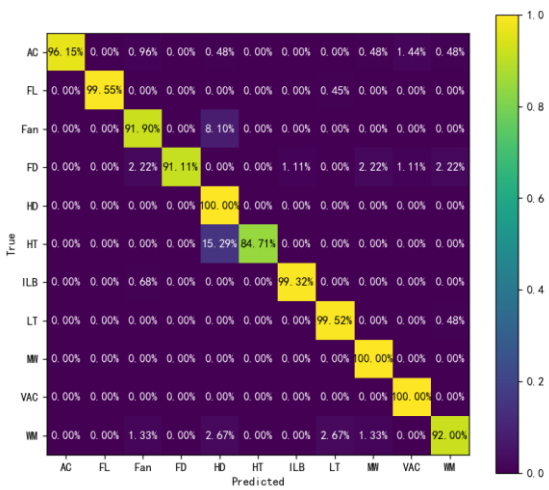


FIGURE 3. Confusion Matrix of SESR.

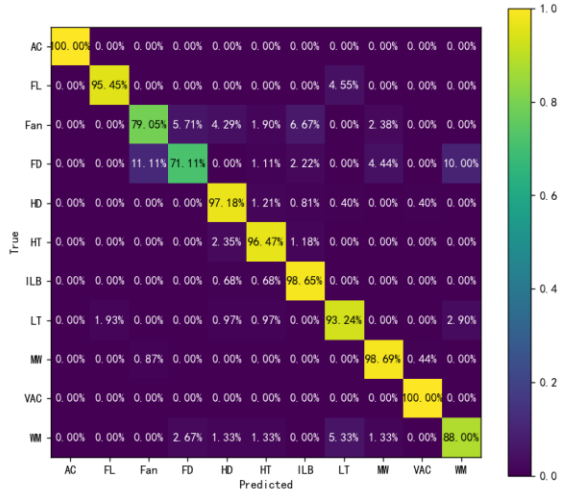


FIGURE 4. Confusion Matrix of Mean Teacher.

From Fig. 3 we can observe that most of the appliances show good performance, especially when the true labels are hair dryer, microwave, and vacuum cleaner, all of which are identified, reaching 100%. The performance of the heater is poor, with 15.29% of the heater being incorrectly identified as hairdryer. For fan, there are 8.10% of fan were incorrectly identified as hairdryer. It can be seen that although the recall rate of hairdryer has reached 100%, its accuracy is not high.

As can be seen from Fig. 4, for the Mean Teacher method, the performance of fan and refrigerator is poor, and it is easy to confuse them. When the true label is fan, 5.71% is wrongly predicted as refrigerator. When the true label is the refrigerator, 11.11% are incorrectly predicted as fan.

C. The semi-supervised learning results (using different rates of labels)

The F_{score} results of the model when the proposed method is trained with different label rates are shown in TABLE 1.

TABLE 1
RESULTS OF SESR USING VARIOUS APPLIANCES AT DIFFERENT LABEL RATES ON PLAID

Appliances	F_{score}			
	All labels	*550 labels*	*330 labels*	*110 labels*
AC	0.980	0.925	0.921	0.597
CFL	0.997	0.983	0.964	0.868
Fan	0.943	0.914	0.892	0.619
FD	0.953	0.804	0.811	0.579
HD	0.938	0.950	0.885	0.706
HT	0.917	0.924	0.839	0.726
ILB	0.993	0.95	0.911	0.832
LT	0.990	0.941	0.892	0.763
MW	0.991	0.982	0.973	0.903
VAC	0.973	0.961	0.972	0.883
WM	0.932	0.874	0.789	0.753

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

It can be found that when all the data were labeled, each appliance achieved good results. The F_{score} of fluorescent lamp, incandescent light bulb, microwave oven was above 0.99, and other appliances also reached 0.9. When unlabeled data was added to the training data, the F_{score} results were not significantly affected. For the microwave, the F_{score} dropped by only 0.088 from all labeled data to only 110 labeled data, and with only 110 labels the F_{score} was still 0.903, which is satisfactory. Secondly, vacuum cleaner, fluorescent lamp and incandescent light bulb are not greatly affected. The appliances that are more affected are air conditioner, fan, and refrigerator. For fan, the F_{score} was still 0.892 when the data used 330 labels, but when the label rate then dropped to 110 labels, the F_{score} dropped straight to 0.619. For fridge, a sharp drop in performance occurs too when using 110 labels, straight down to 0.579. The air conditioner even dropped from 0.98 to 0.597 with the change in label rate. The self-supervised method proposed in this paper affects all these three appliances when using a large amount of unlabeled data. Overall, each appliance can perform satisfactorily when the control label rate is not very low.

For the semi-supervised performance comparison, we have chosen two methods, Mean Teacher [28] and CoMatch [13], as comparison terms. Mean Teacher has been introduced in the supervised learning results (using all labels). CoMatch is a semi-supervised learning method based on contrastive learning. It jointly learns two representations of the training data: class probabilities and low-dimensional embeddings. In this method, given a batch of unlabeled images, their weakly augmented images are used to generate memory-smoothed pseudo-labels that are used as targets for training class predictions on strongly augmented images. A pseudo-labeled graph is constructed to measure the similarity between samples, which is used to train the embedding graph such that images with similar pseudo-labels have similar embeddings. The differences between the proposed SESR method and Mean Teacher and CoMatch at different label rates are reflected in the following Fig. 5 - Fig. 7.

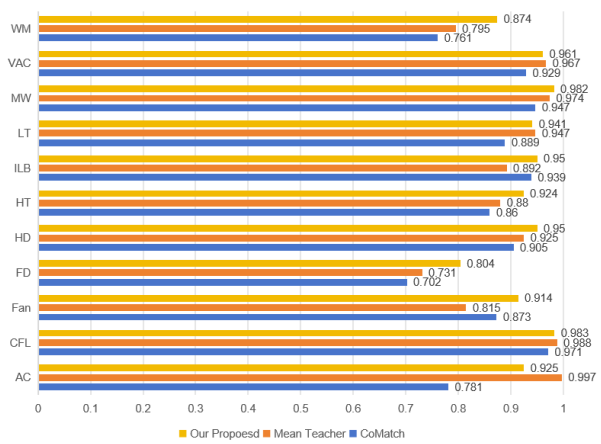


FIGURE 5. Comparison results using 550 labels on PLAID dataset.

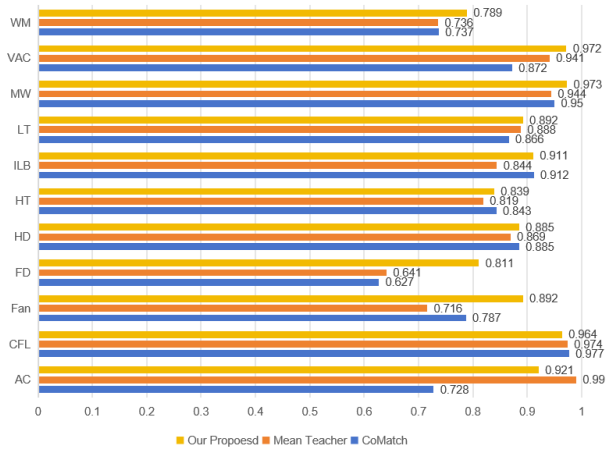


FIGURE 6. Comparison results using 330 labels on PLAID dataset.

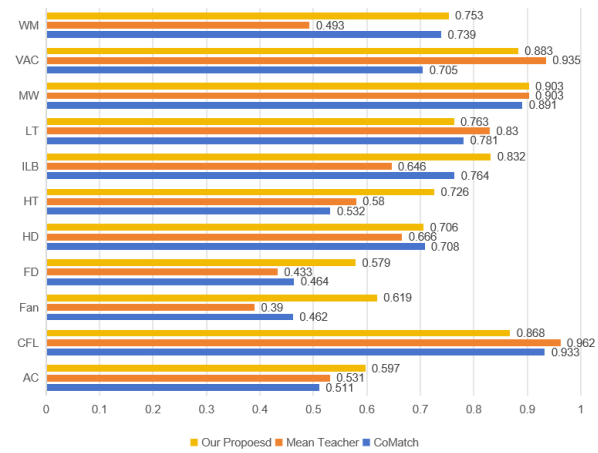


FIGURE 7. Comparison results using 110 labels on PLAID dataset

We can see that the proposed method performs better than Mean Teacher for most appliances, except for vacuum cleaner and air conditioner, which F_{score} are slightly below the Mean Teacher method. Compared with Mean Teacher method, our method has a strong advantage in the fridge, washing machine, fluorescent lamp, and other electrical appliances, especially when the label data is gradually reduced, the advantage is more obvious. Compared with CoMatch, the F_{score} of several appliances： microwave, incandescent light bulb, hairdryers, and compact fluorescent lamp, do not differ much between these two methods. For other appliances besides, the F_{score} of our proposed method outperforms CoMatch, especially in fridge.

F_{score} is used to represent the accuracy of each appliance under different label rates, but the overall performance of each method cannot be intuitively seen. Therefore, F_{macro} is further compared and analyzed, and the results are shown in Table 2.

It can be seen that F_{macro} of the proposed SESR method is higher than that of the Mean Teacher method under different label rates, which further verifies the accuracy of this method.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

TABLE 2

THE RESULTS OF F_{MACRO}

Method	F_{macro}			
	All labels	*550 labels*	*330 labels*	*110 labels*
Proposed SESR	0.965	0.928	0.896	0.748
Mean Teacher	0.920	0.901	0.852	0.707
CoMatch		0.928	0.895	0.748

We can see that the F_{macro} of our method is 0.965 when using all labeled data, 0.928 when using 550 labels and 0.896 when using 330 labels, with very small decrease in F_{macro}. Therefore, it can be considered that our method is more adapted to the case where there is a moderate amount of labeled data and a large amount of unlabeled data in the training data. When there are few labeled data, although F_{macro} performance is better than Mean Teacher, it does not achieve the expected performance. The proposed method is basically equal to F_{macro} of CoMatch, but considering the F_{score} together, our method is more competitive.

We speculate that the reasons for the poor performance of SESR at low label rates are as follows: Gidaris et al. (2018)[21] showed that their self-supervised model focused attention mapping on the salient part of the image to help complete the rotation identification task. In real experiments, however, the middle number (the number to be identified) is usually surrounded by "jamming" numbers. As the numbers are rotated and flipped, the self-supervised branch may pick up the dominant visual features corresponding to the jamming numbers and associate them with the supervised branch. These "errors" are most prominent when few labels exist. However, when the label rate is high, the supervised branch is able to correct errors and our SESR model can generate the better identification results.

Except from the confusion matrix, F_{score} and F_{macro}, we can see that our method is better than Mean Teacher and CoMatch. In addition, because the Mean Teacher network is composed of both the teacher network and the student network, Mean Teacher has a large number of hyperparameters to be adjusted, and our method has fewer hyperparameters.

IV. conclusion

A simple but effective multi-tasking CNN framework for self-supervised regularization of supervised and semi-supervised learning is presented in this paper. Our method uses unlabeled data to generate proxy labels, which are combined with true labels to train the model together. The usage of labeled data is reduced and the utilization of unlabeled data is enhanced. The models are respectively supervised branch and self-supervised branch. The two branches move forward together and share a CNN parameter to generate supervised loss and self-supervised loss respectively. Experimental tests were conducted on the PLAID public data set, compared with the Mean Teacher method. When the data are all labeled, the F_{macro} result of SESR is 0.965, which is much higher than the 0.92 of mean teacher. When utilized unlabeled data, the F_{macro} of the proposed method decreases more slowly as the labeling rate decreases, and the F_{score} of each appliance at different label rates shows that the proposed method outperforms the Mean Teacher and CoMatch in most cases. In addition, compared with the traditional self-supervision, our model does not need a separate pre-training step, and does not need to introduce extra hyperparameters for tuning compared with the consistency regularization method.

REFERENCE

[1] G. W. Hart, "Nonintrusive appliance load monitoring", Proc. IEEE, vol. 80, no. 12, pp. 1870-1870, Dec. 1992.
[2] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions", Energy Buildings, vol. 116, pp. 96-103, Mar. 2016.
[3] R. Zachar, P. Lindahl, J. Donnal, W. Cotta, C. Schantz and S. B. Leeb, "Utilizing spin-down transients for vibration-based diagnostics of resiliently mounted machines", IEEE Trans. Instrum. Meas, vol. 65, no. 7, pp. 1641-1650, Jul. 2016.
[4] W. Li, A. Monti and F. Ponci, "Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks", IEEE Trans. Instrum. Meas, vol. 63, no. 11, pp. 2651-2665, Nov. 2014.
[5] N. F. Esa, M. P. Abdullah and M. Y. Hassan, "A review disaggregation method in non-intrusive appliance load monitoring", Renew. Sustain. Energy Rev, vol. 66, pp. 163-173, Dec. 2016.
[6] M. Z. A. Bhotto, S. Makonin and I. V. Baji', "Load disaggregation based on aided linear integer programming", IEEE Trans. Circuits Syst. II Exp. Briefs, vol. 64, no. 7, pp. 792-796, Jul. 2017.
[7] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions", Energy Buildings, vol. 116, pp. 96-103, Mar. 2016.
[8] Shin, Changho, et al. "Data requirements for applying machine learning to energy disaggregation." Energies 12.9 (2019): 1696.
[9] Moradi, Mohammadamin, Yang Weng, and Ying-Cheng Lai. "Defending Smart Electrical Power Grids against Cyberattacks with Deep Q-Learning." P R X Energy 1 (2022): 033005.
[10] Mosavi, Amir, et al. "State of the art of machine learning models in energy systems, a systematic review." Energies 12.7 (2019): 1301.
[11] T. Chen, S. Kornblith, M. Norouzi and G. Hinton, "A simple framework for contrastive learning of visual representations", Proc. Int. Conf. Mach. Learn., pp. 1597-1607, 2020.
[12] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Graph contrastive learning with adaptive augmentation", Proc. Web Conf., pp. 2069-2080, 2021.
[13] J. Li, C. Xiong, and S. C. Hoi, "CoMatch: Semi-supervised learning with contrastive graph regularization," in Proc. IEEE/CVF Int. Conf.Comput. Vis., Oct. 2021, pp. 9475–9484.
[14] L. Wang, B. X. Chen, C. G. Wang, and D. Hua, "Non-intrusive load monitoring algorithm based on features of V–I trajectory", Elect. Power Syst. Res., vol. 157, pp. 134-144, Apr. 2018, [online] Available: http://www.sciencedirect.com/science/article/pii/S0378779617304911.
[15] Y. Liu, X. Wang and W. You, "Non-Intrusive Load Monitoring by Voltage–Current Trajectory Enabled Transfer Learning," in IEEE Transactions on Smart Grid, vol. 10, no. 5, pp. 5609-5619, Sept. 2019.
[16] Faustine, L. Pereira and C. Klemenjak, "Adaptive Weighted Recurrence Graphs for Appliance Recognition in Non-Intrusive Load Monitoring," in IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 398-406, Jan. 2021.
[17] G. E. Hinton, "Reducing the dimensionality of data with neural networks," Science, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[18] D. Pathak et al., "Context encoders: Feature learning by inpainting," in Proc. Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2536–2544.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3337385

[19] G. Larsson, M. Maire, and G. Shakhnarovich, "Colorization as a proxy task for visual understanding," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 6874–6883.
[20] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in International Conference on Learning Representations, 2018.
[21] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in Proceedings of ICLR, 2017.
[22] Y. Luo, J. Zhu, M. Li, Y. Ren, and B. Zhang, "Smooth neighbors on teacher graphs for semi-supervised learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8896–8905.
[23] T. Miyato, S. -I. Maeda, M. Koyama and S. Ishii, "Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 8, pp. 1979-1993, 1 Aug. 2019.
[24] S. Park, J. Park, S. Shin, and I. Moon, "Adversarial dropout for supervised and semi-supervised learning," in AAAI. AAAI Press, 2018, pp. 3917–3924.
[25] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc.
[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in Proc. 27th Int. Conf. Mach. Learn., 2010, pp. 807–814.
[27] J. M. Gillis and W. G. Morsi, "Non-Intrusive Load Monitoring Using Semi-Supervised Machine Learning and Wavelet Design," in IEEE Transactions on Smart Grid, vol. 8, no. 6, pp. 2648-2655, Nov. 2017.
[28] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in Proceedings of NIPS, vol. 30, pp. 1195–1204, 2017



Ruifeng Zhao received the Ph.D. degree from Chongqing University, Chongqing, China in 2013. Currently, he is a senior engineer in Guangdong Power Grid Co., Ltd. Power Dispatching and Control Center.
His research interests include operation and control and optimization techniques in power systems, distribution network automation, network information security of power systems.



Jiangang Lu received the M.S. degree from Chongqing University, Chongqing, China in 1996. Currently, he is a professor-level senior engineer in Guangdong Power Grid Co., Ltd. Power Dispatching and Control Center.
His research interests include the operation and control and optimization techniques in power systems, the distribution network automation, network information security of power systems.



Bo Liu received the B.S., M.S., and Ph.D. degrees in electrical engineering from Tianjin University, Tianjin, China, in 2011, 2014, and 2019, respectively. He is the Associate Professor with the School of Electrical Automation and Information Engineering, Tianjin University.
His research interests include non-intrusive load monitoring and disaggregation, big data analytics and applications, AI in power systems, smart grid, and the ubiquitous power Internet of Things.



Zhiwen Yu received the M.S. degree from Harbin Institute of Technology, Harbin, China in 2001. He is working in Guangdong Power Grid Co., Ltd. Power Dispatching and Control Center.
His research interests are mainly in the fields of dispatching automation and power system network security, etc.



Yanru Ren is currently working toward the M.S. degree in electrical engineering with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China.
Her research interests include non-intrusive load monitoring techniques.



Wenjie Zheng received the Ph.D. degree from South China University of Technology, Guangzhou, China in 2010. He is working in Guangdong Power Grid Co., Ltd. Power Dispatching and Control Center.
His research interests include power spot market, renewable energy, automatic generation control (AGC), etc.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4