

RTNILM: A Deep Robust Transfer Neural Network for Practical Application of NILM

Xiaohua Pan, Linhui Ye^{ID}, Deyu Weng^{ID}, Jinyin Chen^{ID}, and Jianwei Yin^{ID}, Senior Member, IEEE

Abstract—Nonintrusive load monitoring (NILM) has emerged as a pivotal technology in energy management, garnering significant attention in both research and engineering communities. Despite its potential, conventional NILM methods often exhibit limitations in addressing critical challenges such as domain shift, new appliance detection, and noise interference, thereby hindering their practical application. To overcome these limitations, we propose meta-learning named deep robust transfer neural network (RTNILM), a novel framework that simultaneously addresses these challenges while significantly enhancing NILM performance in real-world scenarios. The RTNILM framework incorporates several innovative components: first, an optimized deep, wide, and robust network architecture is derived through neural architecture search from source domain data set; second, pretrained with optimized general end-to-end loss to acquire a general appliance recognition ability and enhance the model's robustness; third, further trained with model-agnostic meta-learning strategy to improve the model's generalization on the target domain data set; fourth, by comparing the similarities between features from new appliances and known appliances, achieve new appliance detection. Extensive experimental evaluations across three public datasets and one self-collected dataset demonstrate the superiority of RTNILM in cross-domain recognition, new appliance detection, and noise interference. The average improvement in accuracy of cross domain appliance recognition, new appliance detection, and noise interference compared to other methods exceeded 10%, 20%, and 5%, respectively.

Index Terms—Cross-domain, metric learning, model-agnostic meta-learning (MAML), nonintrusive load monitoring (NILM), robustness.

I. INTRODUCTION

IN RECENT years, amid rapid economic and social development, the global energy demand has shown sustained growth [1]. This not only forces serious challenges to the

Received 27 November 2024; revised 26 March 2025; accepted 2 May 2025. Paper no. TII-24-6312. (Corresponding author: Deyu Weng.)

Xiaohua Pan, Linhui Ye, and Deyu Weng are with the Software Innovation Laboratory, Binjiang Institute of Zhejiang University, Hangzhou 310051, China (e-mail: panxh@zju-bj.com; yelin@zju-bj.com; wengdeyu@zju-bj.com).

Jinyin Chen is with the Institute of Cyberspace Security, College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: chenjinyin@zjut.edu.cn).

Jianwei Yin is with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China (e-mail: zjuyjw@cs.zju.edu.cn).

Digital Object Identifier 10.1109/TII.2025.3574351

1941-0050 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

efficiency and reliability of traditional power grids [2] but also increases CO₂ emission, leading to the intensification of global warming. By decomposing individual appliance power consumption from aggregated electrical data, nonintrusive load monitoring (NILM) enables precise management of residential electricity usage and optimizes energy consumption, thereby reducing CO₂ emissions and contributing to environmental preservation.

In the past few years, various NILM methods have been developed. Converting data into 2-D images [3], [4], feature representation [5], [6], [7], [8], [9], and model design [10], [11], [12] are commonly used techniques in load classification. Chen et al. [5] proposed a scale and context-aware convolutional model for load identification by integrating multiscale and contextual information from electrical data. Guo et al. [6] achieved better load identification results with discrete wavelet transform and active deep learning. Yu et al. [13] extracted multitime-scale current shapelets to recognize appliances based on ensembled classifiers and achieved good classification effects.

More recently, considering the practical application of NILM methods, lightweight [14], [15], [16], [17] and scalable [18], [19] methods have also caught the attention of researchers. Lei et al. [17] proposed a lightweight neural network named NILM-LANN with several tricks to deepen the network while maintain a small number of parameters. The proposed method was successfully deployed in a embedding device and achieved excellent results on three different dataset. Furthermore, Athanasiou et al. [16] designed a novel pretraining deep network compression strategy in the area of NILM based on L1 norm. The proposed method greatly reduced the model parameters with negligible performance degradation.

Unfortunately, there are still some practical issues that need to be fully addressed in the real world.

- 1) *Robustness to power grid noise* [20]: The model's robustness to noise in the power grid is insufficiently discussed, even though noise is common in the grid.
- 2) *Detection of new appliance type* [11]: With a wide variety of appliance types, models may encounter appliances that are not learned during training, causing misclassification in practical applications.
- 3) *Model performance degradation caused by domain shift* [21]: Differences in electrical data for appliances across different brands, users have a distribution gap between source and target domain datasets, leading to reduced model performance on the target domain.

Some of the concerns mentioned above have been addressed in recent studies. For example, Fazio et al. [20] proposed a method to manage the off-state noise and improve model performance. Xiao et al. [11] leveraged a one-class classification model to achieve the detection of new appliances with compound features. Solutions [21], [22] that resort to meta-learning [23] have been proven effective in recent years. Wang et al. [22] first applied model-agnostic meta-learning (MAML) in NILM. Combined with ensemble learning, the model achieved better performance with minimal samples in new tasks. However, trained with cross-entropy loss, the traditional MAML methods are incapable of detecting new appliances. Luo et al. [21] proposed a method for practical NILM via metric-based meta-learning, which improved the model performance across different domains. However, existing methods can only address one or two concerns, and cannot simultaneously solve the three aforementioned problems. It should be noted that robustness, new appliance recognition, and domain transformation are common problems in practical application scenarios that need to be solved. The performance of the above methods in the application process is limited.

To address the above issues simultaneously and improve the effectiveness of NILM in practical applications, a new NILM framework based on neural architecture search (NAS), optimized generalized end-to-end (OGE2E) loss, and meta-learning named deep robust transfer neural network (RTNILM) is proposed for practical NILM applications. The innovation of this framework lies in the following aspects.

- 1) To the best of authors' knowledge, this is the first framework in NILM that enables a model to simultaneously possess robustness, new appliance recognition, and cross-domain capabilities.
- 2) We have enhanced the generalized end-to-end (GE2E) loss function to develop the OGE2E loss function and integrated it with the NAS framework, thereby further strengthening the model's feature extraction capabilities and imbuing it with robustness.
- 3) By combining MAML with metric learning, we have achieved superior cross-domain capabilities compared to existing works.

The specific workflow of RTNILM is as follows: an optimal sequential architecture constructed with predefined CNN blocks is obtained using a differentiable NAS algorithm. Then, pretraining the model on the source domain dataset with OGE2E loss to enlarge the distances between features of different types and enhance model robustness. In the end, to address new appliance detection and domain shift problems, the model is further trained with OGE2E loss and MAML strategy.

The main contributions of this work can be summarized as follows.

- 1) A new NILM framework named RTNILM based on NAS, OGE2E loss, and meta-learning is proposed to deal with the problems of noise interference, new appliance detection, and domain shift simultaneously in practical NILM.
- 2) The proposed method first combines metric learning with MAML training strategy achieves a better performance than regular metric-based meta-learning with only a few fine-tuning steps and samples while is capable of identifying new appliances.

- 3) An OGE2E loss consisting of GE2E loss and spectral norm (SN) regularization is proposed not only to obtain discriminative features but also to force the model under Lipschitz constraint and enhance model robustness.
- 4) Extensive experiments are carried out in this article with four different datasets to prove the effectiveness of the proposed method in cross-domain recognition, new appliance detection, and antinoise ability.

Compared with three SOTA methods in NILM, the superiority of the proposed method is further verified.

The rest of this article is organized as follows. The proposed method and practical NILM framework are elaborated in Section II. The experimental results across four different datasets are demonstrated and analyzed in Section III. Finally, Section IV concludes this article.

II. METHODOLOGY

The concepts of the proposed RTNILM contain the following parts:

- 1) robust architecture learning;
- 2) MAML pertaining and fine-tuning;
- 3) loss function and regularization.

The details are elaborated as follows. The overall framework of RTNILM is demonstrated in Fig. 1.

A. Robust Architecture of RTNILM

The robustness of the models needs to be discussed in NILM, as noise in the power grid is common. Increasing the depth and width of deep learning models is a common approach to enhance the robustness [24]. However, randomly stacking networks can complicate model optimization and increase complexity, which may degrade performance. Thus, the differentiable NAS named DARTS [25] is introduced to learn the model's optimal architecture for robust NILM automatically.

DARTS [25] is currently one of the most practical NAS algorithms. Instead of searching for the optimal combination of network parameters, DARTS searches the optimal network from the predefined network blocks. DARTS aims to select the best-performing combination of these predefined blocks on the validation dataset based on the global network architecture.

The proposed method defines the global architecture based on the sequential connection of CNN blocks, as depicted in Fig. 1 where Blocks 1 to Block n are candidate blocks and represent the whole search space. The predefined blocks in this article are shown in Table I. The internal structures of StackConv and DilConv are standard residual blocks. StackConv refers to the stacked CNN with an expansion factor to increase the network's width. DilConv represents dilated CNN, where the dilation coefficient is set to 2 across different kernel sizes.

The advantages of a sequential CNN architecture used in this article are as follows.

- 1) CNN is faster in training and has better robustness compared to RNN [26].
- 2) CNN is easy to expand in depth and width to achieve better robustness and resist noise interference.
- 3) Sequential structures have fewer parameters and are easier to train compared to parallel structures.

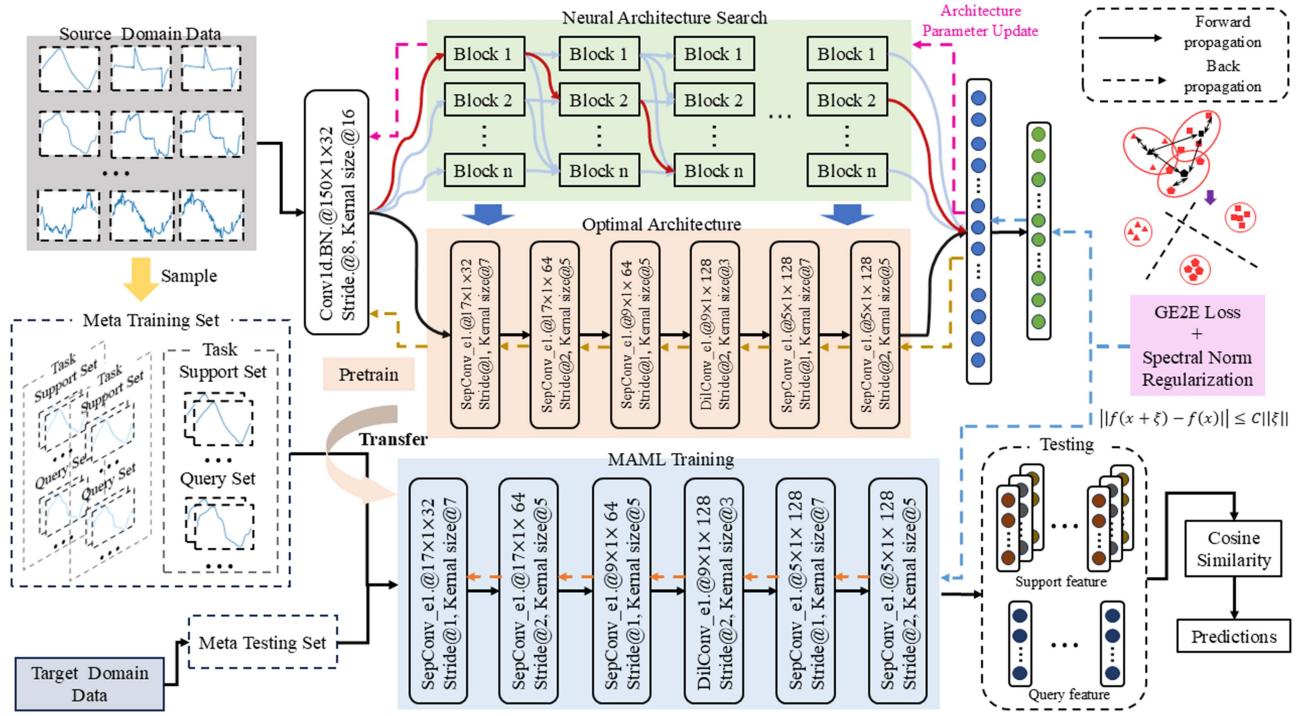


Fig. 1. Illustration of the proposed RTNILM. The proposed NILM framework can be divided into three components: NAS, pretrain, and MAML fine-tuning. At first, the optimal network architecture is obtained with a differentiable NAS. Then, the model is pretrained on a fully supervised source domain dataset. In the end, the pretrained model is fine-tuned with the MAML training strategy. The data preparation and task generation are depicted on the left. On the right, the visualization of OGE2E loss and testing stage are depicted.

TABLE I
PREDEFINED CNN BLOCKS OF RTNILM

Search Block	Kernal Size	Expansion/Dilation
StackConv_3_e1	3	1
StackConv_5_e1	5	1
StackConv_7_e1	7	1
StackConv_3_e2	3	2
StackConv_5_e2	5	2
StackConv_7_e2	7	2
DilConv_3	3	2
DilConv_5	5	2

The specific search process of DARTS can be represented as follows. Suppose S is the search space containing predefined network blocks. Therefore, the output of each layer in the overall architecture can be represented as follows:

$$x^{i+1} = \sum_{s \in S} \frac{\exp(\alpha_s^i)}{\sum_{s' \in S} \exp(\alpha_{s'}^i)} s(x^i) \quad (1)$$

where x^i is the input to the i th layer of the network, $s(x^i)$ represents the output of block s in i th layer, α_s^i denotes the architecture parameters of block s in i th layer. x^{i+1} is the output of the weighted sum of each block, where the weights are α_s^i constrained via softmax function. The architecture search process consists of two steps: in the first step, the network parameters w

are updated using training set with fixed architecture parameters. In the second step, the architecture parameters α_s^i are updated based on the loss of the validation set. Once the model converges, the optimal block in each layer is selected based on the largest softmax weights. The objective of the architecture search is to minimize the validation set loss, achieved by updating the network and architecture parameters successively. It can be described as follows:

$$\nabla_w \mathcal{L}_{\text{val}}(w - \epsilon \nabla_{\text{train}}(w, \alpha), \alpha) \quad (2)$$

where \mathcal{L}_{val} is the validation loss, w and α represents the network parameters respectively, ϵ is the learning rate in the first step update. To enhance the efficiency of DARTS, only one step of network parameter update is performed at each epoch.

B. MAML Training in RTNILM

MAML is an optimization-based meta-learning strategy that includes a two-step updating process with inner and outer loops. In the inner loop, the model parameters are updated using support set samples. In the outer loop, the model updates meta-knowledge with query set loss. The update strategy of MAML during the training stage is as follows:

$$\theta^* = \operatorname{argmin}_{\theta} E_{(x,y) \in D_S \sim M_{Tr}} \mathcal{L}(x, y; \theta, w) \quad (3)$$

$$w^* = \operatorname{argmin}_w E_{(x,y) \in D_Q \sim M_{Tr}} \mathcal{L}(x, y; \theta^*(w)) \quad (4)$$

where $M_{Tr} = \{T_1, T_2, \dots, T_3\}$ denotes the meta-training set, which consists of multiple appliance recognition tasks. The

parameter θ represents the task-specific network parameters and w represents the learned meta-knowledge, acting as the initialization parameters for each task. The objective of MAML is to train the model through various tasks and obtain the optimal initialization parameters (meta-knowledge), that allow the model to perform well on new tasks with only several fine-tuning steps on a few samples.

MAML has also been explored in the NILM [22]. However, extensive data is required for training, and new appliance types with cross-entropy loss are unable to be recognized. Different from [22], this article combines metric-learning loss GE2E (introduced in the following section) with the MAML training strategy, improving model performance on different domains while ensuring feature discrimination. Furthermore, trained with GE2E loss, the model is also capable of recognizing new appliances and better aligns with practical NILM applications.

C. OGE2E Loss

To better obtain discriminative and noise-insensitive features, an integrated loss function named OGE2E loss is proposed, which consists of GE2E loss and SN regularization.

1) GE2E Loss: To capture the similarity relationship between a batch of sample data, the GE2E loss [27] is introduced to broaden the distances between different classes and gather samples among the same class. GE2E loss covers the whole batch of the dataset with the maximum usage of data, the loss of a single sample can be represented as

$$\mathcal{L}(e_{ij}) = V_{ij,i} - \log \sum_{k=1}^N \exp(V_{ij,k}) \quad (5)$$

where e_{ij} is the embedding feature of the j th sample in i th class. $V \in \mathbb{R}^{M \times N}$ is the similarity matrix, where M represents the sample number of a single batch, and N is the number of classes. In this way, $V_{ij,i}$ is the scaled cosine similarity between e_{ij} and the centroid of the i th class. The scaled cosine similarity can be represented as

$$V_{ij,k} = w \cos(e_{ij}, c_k) + b \quad (6)$$

where c_k is the centroid of the k th class, w and b are learnable parameters. Thus, the total loss of GE2E can be derived as

$$\mathcal{L}_{\text{GE2E}} = \sum_{i,j} \mathcal{L}(e_{ij}) = \sum_{i,j} V_{ij,i} - \log \sum_{k=1}^N \exp(V_{ij,k}). \quad (7)$$

GE2E loss acts to reduce the distance of samples within the same class and increase the distance of samples from different classes by optimizing a batch of samples rather than pairs. In this way, the cluster of each class is gathered, and the new appliance data will be isolated beyond them. Meanwhile, the separated loss of GE2E generates more discriminative features and improves the model performance of appliance recognition.

2) SN Regularization: SN regularization [28] is a technique that enables deep learning models to approximately achieve the Lipschitz constraint. The condition of the Lipschitz constraint is as follows:

$$\|f(x + \xi) - f(x)\| \leq C\|(x + \xi) - x\| \quad (8)$$

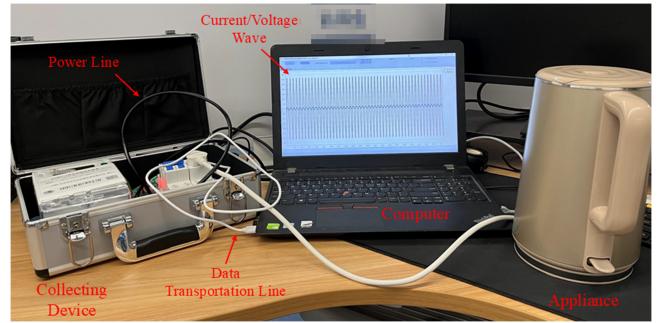


Fig. 2. Data collecting device of the proposed dataset in the laboratory.

where $f(\cdot)$ is the model, C represents the Lipschitz constant that is determined by the model parameters. $\|\cdot\|$ is the L2 norm and ξ is a random noise. The inequality can be interpreted as when the model input is disturbed by a tiny random noise ξ , the L2 norm of perturbation between outputs should be less than or equal to C times the L2 norm of noise ξ . This means the Lipschitz constant C controls the robustness of the model $f(\cdot)$. Without loss of generality, suppose $f(\cdot)$ is a fully connected layer. The 8 can be simplified as

$$\begin{aligned} \frac{\|f(x + \xi) - f(x)\|}{\|\xi\|} &= \frac{\|W(x + \xi) + b - (Wx + b)\|}{\|\xi\|} \\ &= \frac{\|W\xi\|}{\|\xi\|} \leq \sigma(W) = \max_{\xi \neq 0} \frac{\|W\xi\|}{\|\xi\|} \end{aligned} \quad (9)$$

where W and b are parameters of the fully connected layer, $\sigma(W)$ is the SN of matrix W , and controls the robustness of $f(\cdot)$. Therefore, the robustness of the $f(\cdot)$ is controlled by $\sigma(W)$. By constraining the SN of each layer in the network, the model can approximately achieve the Lipschitz constraint and improve the model's robustness.

Thus, the proposed OGE2E loss used can be concluded as GE2E loss plus regularization of the SN

$$\mathcal{L}_{\text{OGE2E}} = \mathcal{L}_{\text{GE2E}} + \lambda \sum_i \sigma(W_i) \quad (10)$$

where λ is a regularization parameter and i refers to the i th operation in the neural network.

D. Conclude the Proposed RTNILM Framework

The overall framework of RTNILM is depicted in Fig. 1. At first, a deeper and wider neural network architecture is obtained via DARTS trained on the source domain dataset and OGE2E loss. A head CNN block in the front and linear layers in the end are frozen for better representations.

After the optimal architecture is achieved, the model is pre-trained with the source domain dataset and OGE2E loss to fully leverage the supervised information. The pretrained model is further learned with the MAML training strategy and meta-training set. The data from tasks are randomly sampled from source domain dataset in this article. The MAML strategy helps the model to perform better in different domains with a few fine-tuning steps and data samples. In training stage, the model

updates the inner loop and outer loop by comparing the cosine similarity between query samples and support samples from different classes. In the testing stage, the model is fine-tuned with support set samples in 5 steps using OGE2E loss. The final labels of query set samples are determined with the highest similarity of support set samples.

By adding SN regularization during the training process, the model approximately satisfies the Lipschitz constraint, reducing sensitivity to small disturbances and noises. Experimental results in Section III prove that it can effectively improve the robustness of the model in NILM.

III. EXPERIMENTS AND ANALYSIS

To validate the effectiveness of the proposed RTNILM, experiments are conducted under the conditions of cross-domain, new appliance detection, and noise interference. The results are compared against other SOTA NILM methods to prove the superiority of the proposed method.

A. Experimental Setup

To increase the network's depth and width, the number of input channels is set to 32 and the number of candidate blocks is 6. The balance ratio λ of SN regularization in OGE2E loss is set to 0.1. To avoid gradient explosion in the process of MAML training, gradient clipping is applied with the clipping norm of 10. The few-shot learning setup N -way K -shots is used in MAML training and testing. To better simulate the scenarios in practical applications, N is set to the size of the whole label space and K is set to 1, 3, and 5 to mimic the small sample scenario. The source domain and target domain datasets are divided into tasks for testing. To reduce randomness, 1000 tasks are randomly selected from the target domain dataset for testing. All experiments use current signals as input data and are resampled to the length of 150.

The comparative methods studied in this article are selected from the NILM area, which is related to our topic.

- 1) *SiameseNet* [29]: SiameseNet is a metric-learning model and trained with contrastive loss in the form of sample pairs.
- 2) *AWRG* [4]: AWRG transforms raw current data into an adaptive two-dimensional recurrence graph and updates the parameters of the graph and model simultaneously during training.
- 3) *FS-RelationNet* [21]: FS-RelationNet is the first application of metric-based meta-learning in NILM and is dedicated to solving the cross-domain problem.

B. Dataset Description

Four different datasets are used in the experiments including three publicly available datasets: PLAID 2018 [30], WHITED [31], and COOLL [32], along with a self-collected dataset.

- 1) The submetered data of the PLAID dataset are used which includes a diverse range of household appliance data,

with 17 types of appliances from 56 brands, totaling 1876 records.

- 2) The WHITED dataset includes appliance data from households and small industrial equipment. It features 56 types of appliances from 131 brands, totaling 1339 records.
- 3) The COOLL dataset was collected from a university laboratory and includes 42 appliances of 12 types, totaling 840 records.
- 4) The dataset proposed in this article consists of 17 common household appliance types, such as washing machines, refrigerators, desktop computers, and hairdryers.

This dataset is collected in the laboratory with the collective device and procedure shown in Fig. 2 under the sampling frequency of 6400 HZ. To enrich the dataset, appliances like washing machines, refrigerators, kettles, rice cookers, and ovens include multiple brands. All data were collected in Hangzhou, China, with a total of 5806 appliance records.

C. Evaluation Metrics

The accuracy and macro $F1$ score ($F1$) are adopted in this article to evaluate the performance of different methods. Accuracy measures the overall predictive performance of the algorithm, and is calculated using the following expression:

$$\text{Acc} = \frac{\text{count}}{N} = \frac{1}{N} \sum_{n=1}^N I(\hat{y}_n = y_n) \quad (11)$$

where N is the total number of data samples in the test dataset, $I(\cdot)$ is the indicator function, which outputs 1 when the input is true and 0 otherwise. \hat{y}_n is the predicted labels and y_n is the ground truth.

The $F1$ score is the harmonic mean of precision and recall, balancing the model's ability to predict and identify specific categories. The precision and recall for the i th class are calculated using the following equations:

$$\text{Precision}_{(i)} = \frac{\text{TP}_{(i)}}{\text{TP}_{(i)} + \text{FP}_{(i)}} \quad (12)$$

$$\text{Recall}_{(i)} = \frac{\text{TP}_{(i)}}{\text{TP}_{(i)} + \text{FN}_{(i)}}. \quad (13)$$

The $F1$ -macro score is commonly used in multiclass tasks and represents the mean of the $F1$ scores across all classes

$$\begin{aligned} F1\text{-macro} &= \frac{1}{C} \sum_{i=1}^K F1_{(i)} \\ &= \frac{1}{C} \sum_{i=1}^K 2 * \frac{\text{Precision}_{(i)} * \text{Recall}_{(i)}}{\text{Precision}_{(i)} + \text{Recall}_{(i)}} \end{aligned} \quad (14)$$

where C represents the number of classes. For binary classification, the $F1$ score is referred to as $F1$ -micro, which corresponds to the $F1$ score for each class.

In our experiments, $F1$ -macro is used to evaluate multiclass classification and $F1$ -micro for binary classification in new appliance detection. In experimental results without given evaluation metrics, $-/-$ represents accuracy/ $F1$.

TABLE II
EXPERIMENTAL RESULTS OF BRAND PERSPECTIVE.

Shots	Methods	PLAID	WHITED	COOLL	Proposed
1-shot	SiameseNetML	0.613/0.595	0.896/0.888	0.573/0.583	0.745/0.710
	Fine-tuned AWRG	0.644/0.574	0.614/0.507	0.625/0.592	0.783/0.741
	FS-RelationNet	0.692/0.664	0.823/0.805	0.822/0.800	0.778/0.744
	RTNILM*	0.645/0.624	0.831/0.816	0.805/0.784	0.774/0.760
3-shots	SiameseNetML	0.708/0.682	0.883/0.878	0.718/0.708	0.786/0.762
	Fine-tuned AWRG	0.750/0.687	0.829/0.765	0.773/0.740	0.840/0.813
	FS-RelationNet	0.682/0.648	0.851/0.841	0.873/0.862	0.790/0.747
	RTNILM*	0.815/0.800	0.926/0.922	0.922/0.919	0.916/0.913
5-shots	SiameseNetML	0.749/0.727	0.879/0.875	0.734/0.724	0.832/0.814
	Fine-tuned AWRG	0.821/0.773	0.887/0.834	0.839/0.809	0.908/0.896
	FS-RelationNet	0.690/0.664	0.874/0.867	0.883/0.878	0.810/0.767
	RTNILM*	0.867/0.857	0.948/0.946	0.943/0.941	0.948/0.946

The bold values indicates the best score.

D. Cross Domain Recognition

Following the experimental setup in [21], this experiment evaluates methods from brands and dataset perspectives. In practical applications, appliances of the same type may come from various brands leading to differences in modality and power in the same appliance type. Furthermore, the transfer between different datasets is also experimented to demonstrate the superiority of RTNILM.

From the perspective of brands, one brand data of each appliance is sampled as the target domain dataset and the remaining data are combined as the source domain dataset. The setup is the same in the dataset perspective. It should be noted that the source and target domain datasets in this section share the same label space, the problem of label heterogeneity is discussed in the next section. As for task generation, all tasks are randomly selected from their respective datasets. Data from the same appliance in each task may come from different brands and datasets in training, corresponding to the most challenging generation technique described in [21]. To avoid randomness of the results, each experiment is repeated 10 times.

To fairly comparison with comparative methods, SiameseNet is trained with meta-learning tasks to improve its performance on different domains. The AWRG is first trained on the source domain dataset using cross-entropy loss and fine-tunes the final fully connected layer in the testing stage as shown in [21].

1) Cross Domain Experiment in Brand Perspective: The experimental results of brand perspective are demonstrated in Table II. It can be seen that when K is equal to 1, RTNILM does not achieve the best results for the extreme lack of data in OGE2E training. However, when K is greater than 1, there is a significant steeper improvement in the performance compared to other methods. With $K = 3$ and 5, the proposed algorithm achieves an accuracy improvement of approximately 6%–11% over the second-best results proving its superiority. RTNILM significantly enhances the performance of cross-brand appliance recognition and is effective in real-world scenarios. Both the meta-learning based SiameseNet and FS-RelationNet trained on pair-wise metric loss do not fully utilize the dataset, resulting in limited performance. RTNILM achieves much better performance with only 5 steps of fine-tuning, compared to 30 fine-tuning steps required by the fine-tuned AWRG. The results

TABLE III
EXPERIMENTAL RESULTS OF DATASET PERSPECTIVE.

Shots	Methods	PLAID	WHITED	COOLL	Proposed
1-shot	SiameseNetML	0.371/0.343	0.629/0.604	0.448/0.429	0.738/0.705
	Fine-tuned AWRG	0.278/0.202	0.255/0.217	0.375/0.184	0.646/0.518
	FS-RelationNet	0.152/0.110	0.305/0.249	0.416/0.405	0.491/0.396
	RTNILM*	0.330/0.320	0.677/0.665	0.408/0.399	0.632/0.606
3-shots	SiameseNetML	0.430/0.407	0.654/0.620	0.544/0.525	0.740/0.702
	Fine-tuned AWRG	0.326/0.258	0.302/0.262	0.560/0.347	0.709/0.608
	FS-RelationNet	0.320/0.285	0.375/0.296	0.433/0.379	0.572/0.472
	RTNILM*	0.472/0.460	0.847/0.841	0.610/0.602	0.785/0.772
5-shots	SiameseNetML	0.475/0.452	0.711/0.683	0.574/0.557	0.794/0.774
	Fine-tuned AWRG	0.448/0.394	0.327/0.278	0.604/0.411	0.722/0.632
	FS-RelationNet	0.366/0.324	0.342/0.246	0.469/0.447	0.717/0.676
	RTNILM*	0.552/0.542	0.907/0.903	0.705/0.696	0.837/0.828

The bold values indicates the best score.

further demonstrate the effectiveness of the proposed OGE2E loss and MAML training strategy in NILM cross-domain applications.

2) Cross Domain Experiment in Dataset Perspective: The results of cross-dataset experiments are demonstrated in Table III. Similar to the results of the cross-brand experiment, RTNILM achieves the best performance in cross-dataset experiments when shots are greater than 1. Meanwhile, the results of RTNILM outperform the second-best methods by approximately 10% over four datasets. Due to the diverse modalities across different brands of the PLAID dataset, the transfer performance in PLAID is lower compared to other datasets. Both the FS-RelationNet and fine-tuned AWRG performed poorly in four testing datasets, especially on the PLAID and WHITED datasets because of diverse brands. Fine-tuned AWRG is hard to achieve a good result with only a few samples of fine-tuning. FS-RelationNet that without a particular training strategy is hard to transfer learned relation metrics to different datasets.

E. New Appliance Detection and Recognition

In the practical application of NILM, the model needs to identify and recognize new appliances which means a heterogeneous label space between the source domain and target domain dataset. In this section, the problem is divided into detection (binary classification) and recognition (multiclass classification). For simplicity, the above four datasets are combined to generate two datasets from different perspectives. 1) Appliance types dataset (AT dataset): considering each appliance type as a class which includes a total of 38 appliance types (one class containing different brands). 2) Individual appliance dataset (IA dataset): considering each appliance as a class includes a total of 218 appliances (appliances from different brands are different classes). The IA dataset is expanded with a random sliding window to enrich the dataset. As for the IA dataset, the model is trained to recognize different appliances which is more suitable to real-life NILM.

Both the label spaces of these two datasets are randomly divided into seen labels and unseen labels with a proportion of 70% and 30% of the entire label space, respectively. The data from seen labels and unseen labels are also denoted as close-set and open-set. The source domain dataset contains close-set data

TABLE IV
EXPERIMENTAL RESULTS OF NEW APPLIANCE DETECTION.

Shots	Dataset	Methods	Siamese NetML	FS-RelationNet	RTNILM*
1-shot	AT dataset	Close	0.278/0.430	0.420/0.588	0.852/0.919
		Open	0.323/0.486	0.474/0.637	0.826/0.904
	IA dataset	Close	0.432/0.598	0.581/0.732	0.943/0.970
		Open	0.589/0.737	0.561/0.715	0.843/0.913
3-shots	AT dataset	Close	0.272/0.422	0.454/0.621	0.914/0.955
		Open	0.323/0.486	0.545/0.704	0.845/0.914
	IA dataset	Close	0.473/0.638	0.684/0.809	0.858/0.922
		Open	0.546/0.702	0.472/0.637	0.855/0.921
5-shots	AT dataset	Close	0.309/0.465	0.522/0.683	0.917/0.956
		Open	0.303/0.462	0.606/0.753	0.852/0.919
	IA dataset	Close	0.445/0.611	0.523/0.683	0.912/0.953
		Open	0.597/0.744	0.514/0.675	0.915/0.955

The bold values indicates the best score.

only and the target domain dataset contains both close-set and open-set data with an equal proportion. In the experiment of new appliance detection, the model is trained with source domain dataset and the trained embeddings are recorded. Compare similarities between recorded embeddings and embeddings from the target domain dataset. If the biggest similarity is greater than the predefined threshold, it is classified as seen appliance otherwise it is new. The new appliance recognition experiment further identifies the specific class of the detected seen appliance. Only the correctly detected data in the binary experiment are further recognized in the multiclass experiment. Meanwhile, the model performance of close-set and open-set in the target domain dataset is recorded separately to better visualize the model performance.

This article divides the new appliance detection problem into two experiments due to the practical considerations of NILM applications: 1) It is common for each household to acquire new appliances over time. The binary classification experiment thus evaluates the algorithm's capability to actively discover new appliances, which facilitates user interaction and management of these new devices. 2) The multiclass classification experiment follows the binary classification. Here, the model not only distinguishes new appliances but also identifies them. After recognizing a new appliance, the model actively records its feature by storing it in the feature repository. This allows the model to perform multiclass classification of new appliance types by extracting deep features from the model.

1) Experiments of New Appliance Detection: The experiment results of new appliance detection of two datasets are shown in Table IV. In cases where the similarity between close-set and open-set data cannot be effectively distinguished, the accuracy for both sets should be close to 50%. It can be seen that SiameseNet and FS-RelationNet both perform poorly in two datasets which means they are not qualified to distinguish new appliances. Meanwhile, the proposed RTNILM still achieves the best results and the detection accuracy is near 90%, proving the effectiveness of OGE2E loss in feature separation and gathering.

2) Experiments of New Appliance Recognition: The experimental results of new appliance recognition are presented in Table V. It can be seen that the proposed method achieves

TABLE V
EXPERIMENTAL RESULTS OF NEW APPLIANCE RECOGNITION.

Shots	Dataset	Methods	Siamese NetML	FS-RelationNet	RTNILM*
1-shot	AT dataset	Close	0.6444/0.4674	0.7279/0.5298	0.9999/1.0
		Open	0.3900/0.3137	0.5170/0.3723	0.4414/0.2571
	IA dataset	Close	0.0566/0.0323	0.5812/0.4264	0.9954/0.9910
		Open	0.0843/0.0414	0.4987/0.3225	0.8913/0.7833
3-shots	AT dataset	Close	0.6591/0.4702	0.7278/0.5962	0.9932/0.9848
		Open	0.6000/0.3847	0.4378/0.2928	0.6106/0.4771
	IA dataset	Close	0.0686/0.0379	0.5107/0.3694	0.9801/0.9631
		Open	0.0767/0.0373	0.5863/0.5863	0.8856/0.7705
5-shots	AT dataset	Close	0.6700/0.5478	0.8047/0.6911	0.9932/0.9942
		Open	0.5851/0.4656	0.4095/0.2643	0.7954/0.6682
	IA dataset	Close	0.0690/0.0391	0.3394/0.2178	0.9963/0.9928
		Open	0.0766/0.0388	0.2097/0.1313	0.9061/0.8104

The bold values indicates the best score.

the best recognition results compared to SiameseNet and FS-RelationNet. The classification results from the open-set significantly improve as the shots increase. In both datasets, the performance is substantially higher than comparative methods and is close to 100% in close-set. The SiameseNetML is hard to learn discriminative features in the IA dataset because of the huge amount of classes. Meanwhile, compared to SiameseNetML and FS-RelationNet, the correctly detected data size of RTNILM is bigger because of the excellent results in binary experiments. Thus, the accuracy of FS-RelationNet is higher 7% than RTNILM in the 1-shot experiment.

The experimental results in Section III-D prove that RTNILM is capable of detecting and recognizing new appliances under limited data samples, further proving the effectiveness of RTNILM in practical NILM applications.

F. Appliance Recognition Under Noise Interference

In this section, the robustness of the RTNILM model is evaluated under noisy conditions. In real-world power grids, the collected electrical data often contain significant noise, which can affect the accuracy of appliance recognition. To simulate the impact of noise on the model's performance, this section extends the experiments conducted in Section III-D by introducing three types of noise into the target domain dataset: Gaussian noise, impulse noise, and power grid background noise collected from the power grid when there is no appliances are active. These noises are applied with signal-to-noise ratios (SNRs) ranging from -30 to 50, and the corresponding accuracy is recorded. Since obtaining actual power grid noise data is challenging, Gaussian noise is used to simulate random fluctuations in the grid, impulse noise is employed to mimic sudden current changes, and the collected power grid background noise is utilized to represent real-world noise conditions in the power grid. In this article, the Gaussian noise is generated from a normal distribution with a mean of 0 and a nonfixed variance. The impulse noise is introduced with a ratio of 0.04, and its amplitude varies according to the SNR.

In addition, the contribution of each component in RTNILM is evaluated with an ablation study under noise interference. The IA dataset with more data samples is experimented to better illustrate the accuracy-SNR curves. Meanwhile, the robustness

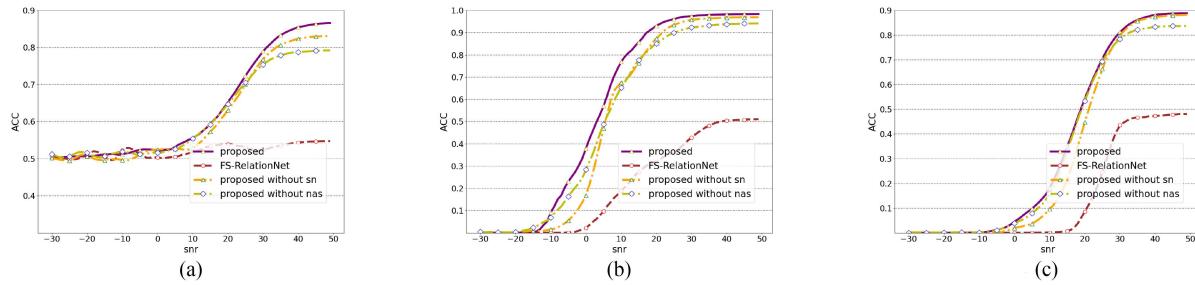


Fig. 3. Robust experiment results of (a) new appliance detection, (b) close-set recognition, and (c) open-set recognition under Gaussian noise.

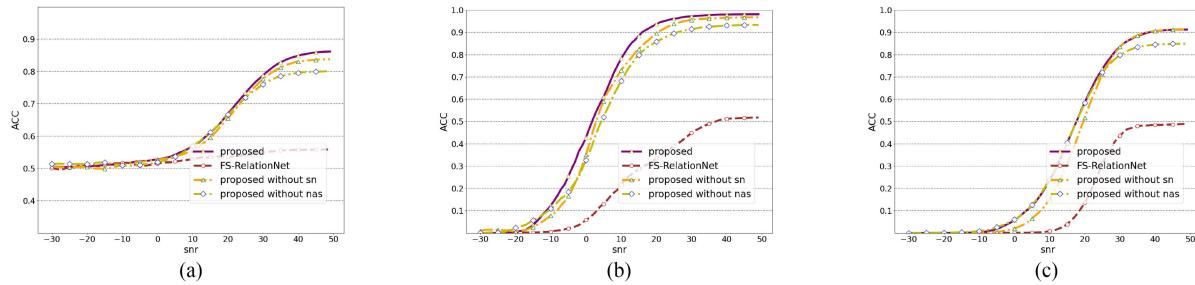


Fig. 4. Robust experiment results of (a) new appliance detection, (b) close-set recognition, and (c) open-set recognition under impulse noise.

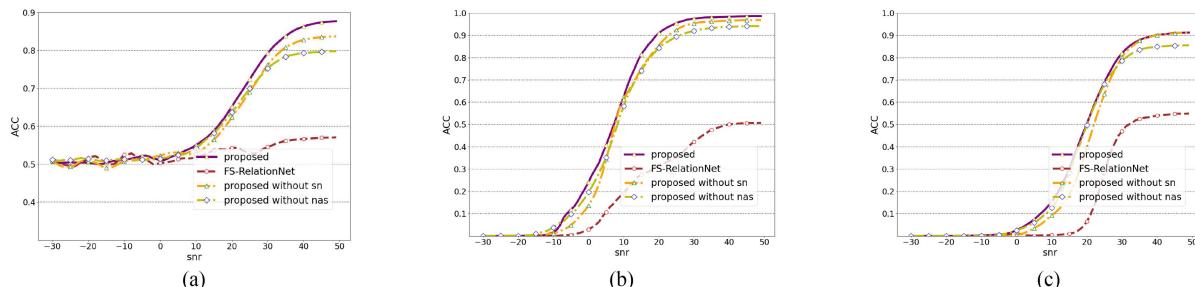


Fig. 5. Robust experiment results of (a) new appliance detection, (b) close-set recognition, and (c) open-set recognition under power grid background noise.

of cross-domain performance is also considered because of new appliance recognition. The results of appliance detection and appliance recognition of close-set and open-set are recorded.

1) Experiment of Robustness and Ablation Study: The accuracy-SNR curves of RTNILM, FS-RelationNet, RTNILM without sn (SN regularization), and RTNILM without NAS of three different noises are shown in Figs. 3–5, respectively. From the overall results, it can be observed that RTNILM significantly outperforms FS-RelationNet under various types of noise and different SNRs. Under different noise interferences, RTNILM demonstrates its ability to resist noise and achieve rising accuracy in both close-set recognition and open-set recognition at SNRs of -20 dB and -10 dB. In contrast, FS-RelationNet only begins to show performance improvement at SNRs of 0 dB and 10 dB. In addition, the SNR at which RTNILM's performance starts to converge is much lower than that of FS-RelationNet, further proving the effectiveness and superiority of RTNILM in combating noise. Through comparisons with RTNILM under different types of noise and SNRs, it is further demonstrated that the OGE2E metric loss adopted by RTNILM can better

aggregate samples of the same class, resulting in model features that are less sensitive to noise. In the ablation study of NAS, the model architecture is constructed with human experiences via the predefined blocks. Comparing the results of RTNILM and RTNILM without sn and NAS, the curves are moving to the large side of SNR in the ablation study and the final results are lower than RTNILM. It proves the effectiveness of SN regularization and learned optimal architecture in counteracting noise and appliance recognition.

Besides, a toy noisy dataset collected in a real household from a power grid is used to quantify the performance of robustness. Different from the collective device in the laboratory, the data acquisition equipment is connected to the power bus directly. Thus, the collected data in the household is more inclined to be interfered by noises in the power system. The real collected toy noisy dataset includes household appliances Hood, washing machine, water heater, water kettle, and water dispenser. The comparative example of Hood in the noisy dataset and the data collected in the lab is shown in Fig. 6. It can be seen that there is a lot of high-frequency noise in real residential data

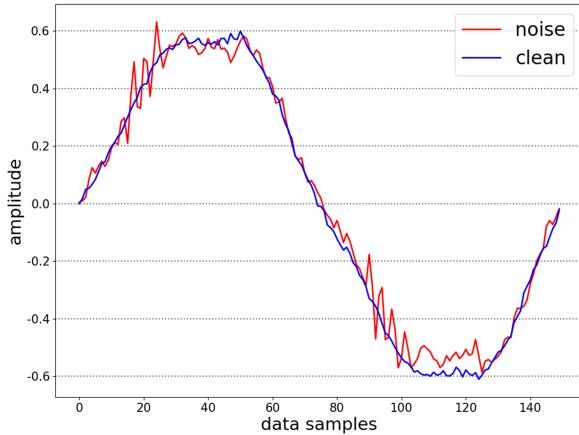


Fig. 6. Comparison of the real residential power grid and proposed laboratory collected data of Hood.

TABLE VI
EXPERIMENTAL RESULTS OF THE REAL COLLECTED NOISY DATASET.

Methods	Metrics	1-shot	3-shots	5-shots
SiameseNet	Acc	0.5991	0.7631	0.7815
	F1-macro	0.5242	0.7286	0.7615
FS-RelationNet	Acc	0.5271	0.7487	0.7407
	F1-macro	0.4290	0.7258	0.7150
* RTNILM	Acc	0.5532	0.7908	0.8168
	F1-macro	0.5363	0.7825	0.8099

The bold values indicates the best score.

compared to the lab data. The noisy dataset recognition results are demonstrated in Table VI (the models are from Section III-E of the IA dataset). It can be seen that RTNILM achieves the best performance in both accuracy and *F1* score. The experiment of real collected noisy data further proves the robustness of RTNILM in practical NILM applications. The collective device illustration in the household is demonstrated in Fig. 7.

2) Visualization and Analysis of Model's Robustness: In this section, the robustness and performance of the model is further visualized. Specifically, data from 5 classes in the close-set and 3 classes in the open-set is randomly selected for data distribution visualization. To better demonstrate the distribution of data in high-dimensional space, this article projects data into 3-D using the tsne3D algorithm and visualizes data on a spherical surface. The visualization results of both close-set and open-set data are demonstrated in Fig. 8. It can be seen from the results that without noise, different classes of data features have better discriminability in RTNILM, which proves the effectiveness of OGE2E loss in obtaining discriminative features. FS-RelationNet trained with mean square error resulting in an overlapped data distribution and poor recognition performance.

To better demonstrate the ability of methods to resist noise, noise of 30 dB SNR is added and visualized together with clean data. It can be seen that with the noise of 30 dB, the noisy data are outside the original distribution in FS-RelationNet, which is consistent with the robustness results under different noises. While the noise data and the original data are distributed in the same location in RTNILM, further proves the robustness and superiority of the proposed algorithm against FS-RelationNet.

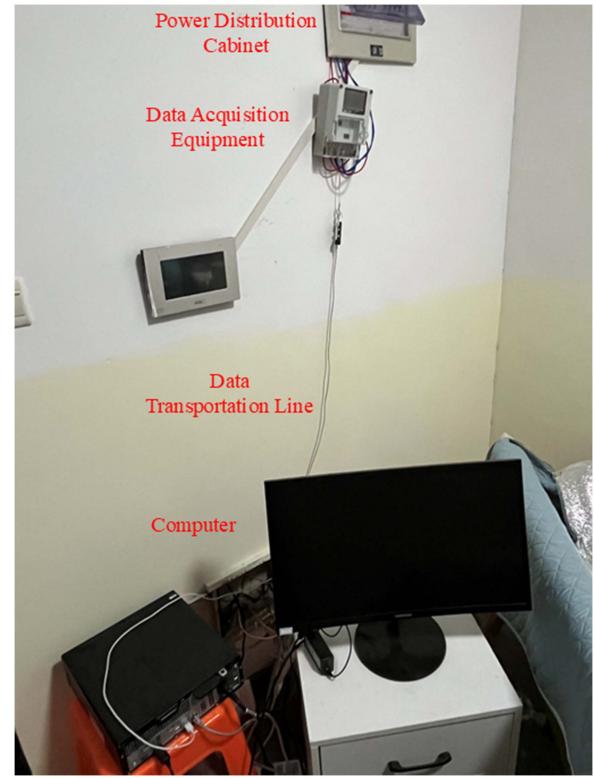


Fig. 7. Data collecting device from the power line bus in the real household. Household electrical data is interfered with by background noise.

G. Experiments of Parameter Sensitivity

To better demonstrate the influence of different hyperparameters on the experimental results, the sensitivity experiments are conducted on parameters including the learning rate in the MAML inner loop training (denoted as task lr), the learning rate in the MAML outer loop training (denoted as meta lr), and the SN regularization parameter λ . The results are shown in Fig. 9.

To investigate the influence of task lr and meta lr on model performance, the values are set from $1e-3$ to $10e-2$ and $1e-5$ and $10e-4$, respectively, recording the accuracy in new appliance detection and appliance recognition on the IA dataset. The results of task lr and meta lr have the same trend of change. The experimental results indicate that when the task lr is from $1e-3$ to $1e-2$ and the meta lr is from $1e-5$ to $10e-4$, the model's performance remains stable and at its optimal level. However, when the learning rates continue to increase, the model's performance significantly declines and starts to fluctuate. This indicates that the MAML training process is highly sensitive to the learning rate; excessively high learning rates can cause the model to deviate from the optimal range, leading to performance fluctuations. Therefore, in practical applications, to ensure stable training, the task lr and meta lr should be set to a relatively small value, and the number of epochs can be appropriately increased to maintain the stability of model training.

To investigate the impact of the SN regularization parameter on model performance, we vary it from 0.01 to 1, add noise with a fixed SNR of 20 dB to the IA data set, and record accuracy on both clean and noisy IA datasets. The experimental results

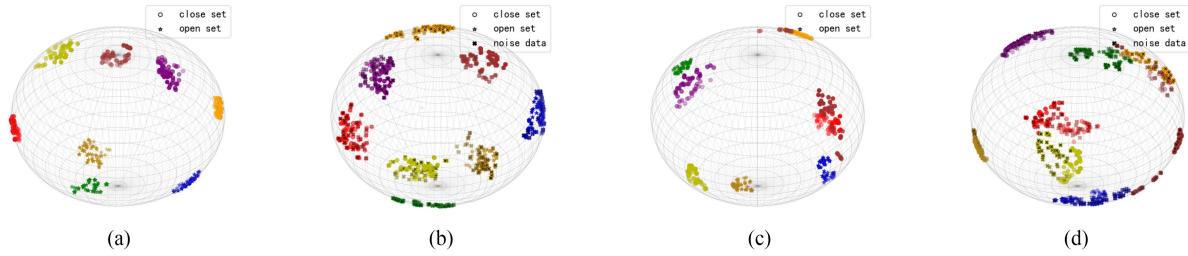


Fig. 8. Visualization of embedding features in 3-D use of (a) RTNILM no noise, (b) RTNILM with noise, (c) FS-RelationNet no noise, and (d) FS-RelationNet with noise.

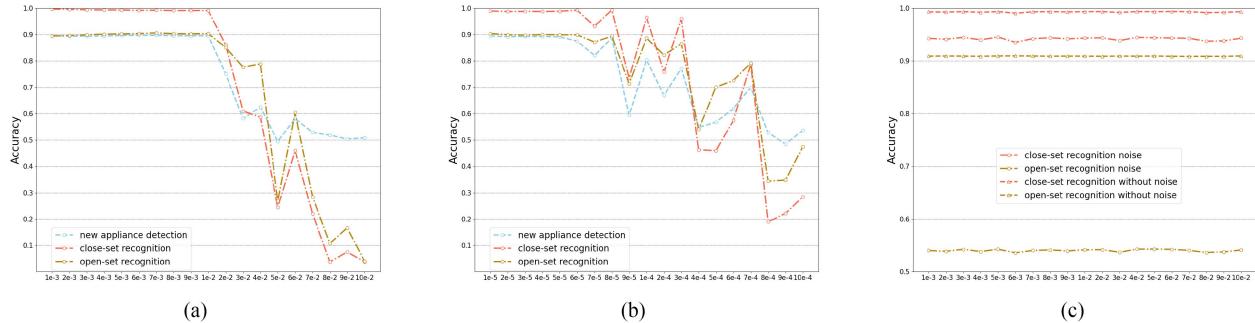


Fig. 9. Sensitivity experiments of different parameters. (a) Task lr. (b) Meta lr. (c) Parameter of SN regularization.

indicate that the model's performance remains relatively stable on both noisy and clean datasets as the spectral regularization parameter changes. This stability is attributed to the use of the GE2E metric loss during model training. It can be seen from (7), that the second part of GE2E loss is log of the sum of multiple small values, the derivative of the second part of GE2E loss is the reciprocal of a small value. Therefore, in the process of backpropagation, gradient explosion is easy to occur. To prevent gradient explosion, a gradient clipping strategy was employed in both the pretraining and MAML training stage (with a gradient clipping coefficient of 3 during pretraining and 10 during MAML training). Therefore, the gradient values updated at each step are close across different SN regularization parameters, leading to stable model performance. Within this gradient clipping coefficient, the SN regularization parameter can be selected from 0.01 to 1 and achieve good model performance.

IV. CONCLUSION

This article presents an innovative solution for practical NILM named RTNILM to tackle the problems of robustness, new appliance detection, and cross-domain. At first, the optimal components of a deep and wide neural network are obtained using DARTS. The model is pretrained to fully leverage the source domain dataset with OGE2E loss. After that, the MAML strategy is employed to fine-tune and equip the model with the ability of cross-domain recognition. Extensive experiments across four datasets and three comparative methods proved the effectiveness and superiority of the proposed method in practical NILM.

In the future, explorations toward one-shot learning will be made to improve the appliance recognition performance if only

one data record is available. Techniques that stabilize the training process of meta-learning will also be studied. Meanwhile, the best combination of gradient clipping coefficients and SN regularization parameters will be explored.

REFERENCES

- [1] “National statistics bureau of China-the electrical consumption in recent years,” [Online]. Available: <https://data.stats.gov.cn/easyquery.htm?cn=C01>
- [2] L. Yan, W. Tian, J. Han, and Z. Li, “Event-driven two-stage solution to non-intrusive load monitoring,” *Appl. Energy*, vol. 311, 2022, Art. no. 118627.
- [3] J. Chen, X. Wang, X. Zhang, and W. Zhang, “Temporal and spectral feature learning with two-stream convolutional neural networks for appliance recognition in NILM,” *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 762–772, Jan. 2022.
- [4] A. Faustine, L. Pereira, and C. Klemenjak, “Adaptive weighted recurrence graphs for appliance recognition in non-intrusive load monitoring,” *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 398–406, Jan. 2021.
- [5] K. Chen, Y. Zhang, Q. Wang, J. Hu, H. Fan, and J. He, “Scale-and context-aware convolutional non-intrusive load monitoring,” *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2362–2373, May 2020.
- [6] L. Guo, S. Wang, H. Chen, and Q. Shi, “A load identification method based on active deep learning and discrete wavelet transform,” *IEEE Access*, vol. 8, pp. 113932–113942, 2020.
- [7] E. Tabanelli, D. Brunelli, A. Acquaviva, and L. Benini, “Trimming feature extraction and inference for MCU-based edge NILM: A systematic approach,” *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 943–952, Feb. 2022.
- [8] Y. Liu, L. Bai, J. Ma, W. Wang, and W. Ouyang, “Self-supervised feature learning for appliance recognition in nonintrusive load monitoring,” *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1698–1710, Feb. 2024.
- [9] Y. Han, H. Chen, J. Wu, and Q. Zhao, “Reconstruction-based supervised contrastive learning for unknown device identification in nonintrusive load monitoring,” *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 2511313.
- [10] L. da Silva Nolasco, A. E. Lazzaretti, and B. M. Mulinari, “DeepDFML-NILM: A new CNN-based architecture for detection, feature extraction and multi-label classification in NILM signals,” *IEEE Sensors J.*, vol. 22, no. 1, pp. 501–509, Jan. 2022.

- [11] J. Xiao, M. Tan, R. Pan, Y. Su, T. Li, and Z. Xu, "Non-intrusive load identification considering unknown load based on bi-modal fusion and one-class classification," *IEEE Trans. Instrum. Meas.*, vol. 73, 2023, Art. no. 2504411.
- [12] E. L. de Aguiar, L. da Silva Nolasco, A. E. Lazzaretti, D. R. Pipa, and H. S. Lopes, "ST-NILM: A wavelet scattering-based architecture for feature extraction and multilabel classification in NILM signals," *IEEE Sensors J.*, vol. 24, no. 7, pp. 10540–10550, Apr. 2024.
- [13] H. Yu, C. Xu, G. Geng, and Q. Jiang, "Multi-time-scale shapelet-based feature extraction for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 1116–1128, Jan. 2024.
- [14] Z. Lu, Y. Cheng, M. Zhong, W. Luan, Y. Ye, and G. Wang, "LightNILM: Lightweight neural network methods for non-intrusive load monitoring," in *Proc. ACM Int. Conf. Syst. Energy-Efficient Buildings Cities, Transp.*, 2022, pp. 383–387.
- [15] Y. Li, R. Yao, D. Qin, and Y. Wang, "Lightweight federated learning for on-device non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 16, no. 2, pp. 1950–1961, Mar. 2025.
- [16] S. Athanasoulias, S. Sykiotis, N. Temenos, A. Doulamis, and N. Doulamis, "A pre-training pruning strategy for enabling lightweight non-intrusive load monitoring on edge devices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. Workshops*, 2024, pp. 249–253.
- [17] Y. Lei, Z. Feng, X. Lin, D. Cao, and J. Zhang, "NILM-LANN: A lightweight attention-based neural network in non-intrusive load monitoring," in *Proc. 27th Int. Conf. Comput. Supported Cooperative Work Des.*, 2024, pp. 181–186.
- [18] C. L. Athanasiadis, T. A. Papadopoulos, and D. I. Doukas, "Real-time non-intrusive load monitoring: A light-weight and scalable approach," *Energy Buildings*, vol. 253, 2021, Art. no. 111523.
- [19] G. Tanoni, L. Stankovic, V. Stankovic, S. Squartini, and E. Principi, "Knowledge distillation for scalable nonintrusive load monitoring," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 4710–4721, Mar. 2024.
- [20] P. Fazio, M. Mehic, and M. Voznak, "Load monitoring and appliance recognition using an inexpensive, low frequency, data-to-image, neural network and network mobility approach for domestic IoT systems," *IEEE Internet Things J.*, vol. 11, no. 8, pp. 13961–13979, Apr. 2024.
- [21] Q. Luo, T. Yu, C. Lan, Y. Huang, Z. Wang, and Z. Pan, "A generalizable method for practical non-intrusive load monitoring via metric-based meta-learning," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 1103–1115, Jan. 2024.
- [22] L. Wang, S. Mao, B. M. Wilamowski, and R. M. Nelms, "Pre-trained models for non-intrusive appliance load monitoring," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 56–68, Mar. 2022.
- [23] A. Vettoruzzo, M.-R. Bouguelia, J. Vanschoren, T. Rognvaldsson, and K. Santosh, "Advances and challenges in meta-learning: A technical review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 4763–4779, Jul. 2024.
- [24] Z. Zhang and H. Peng, "Deeper and wider siamese networks for real-time visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4586–4595.
- [25] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," 2018, *arXiv:1806.09055*.
- [26] J. Chen et al., "RobuSTRMC: Robustness interpretable deep neural network for radio modulation classification," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 4, pp. 1218–1240, Aug. 2024.
- [27] L. Wan, Q. Wang, A. Papir, and I. L. Moreno, "Generalized end-to-end loss for speaker verification," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 4879–4883.
- [28] Y. Yoshida and T. Miyato, "Spectral norm regularization for improving the generalizability of deep learning," 2017, *arXiv:1705.10941*.
- [29] M. Yu, B. Wang, L. Lu, Z. Bao, and D. Qi, "Non-intrusive adaptive load identification based on Siamese network," *IEEE Access*, vol. 10, pp. 11564–11573, 2022.
- [30] R. Medicò et al., "A voltage and current measurement dataset for plug load appliance identification in households," *Sci. Data*, vol. 7, no. 1, 2020, Art. no. 49.
- [31] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, "Whited-a worldwide household and industry transient energy data set," in *Proc. 3rd Int. Workshop noN-Intrusive Load Monit.*, 2016, pp. 1–4.
- [32] T. Picon et al., "COOLL: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification," 2016, *arXiv:1611.05803*.



Xiaohua Pan is currently working toward the Ph.D. degree in computer science with the College of Computer Science and Technology, Zhejiang University, Hangzhou, China.

He is currently a Head Researcher with the Software Innovation Laboratory, Binjiang Institute, Zhejiang University. His research interest covers data mining, deep learning, Internet of Things technology, and their applications in industry.



Linhui Ye received the master's degree in electronic information from the Zhejiang University of Technology, Hangzhou, China, in 2022.

He is currently a Researcher with the Binjiang Institute, Zhejiang University, Hangzhou, China. His research interests include deep learning, data mining, digital signal processing, and adversarial attack and defense.



Deyu Weng was born in Zhejiang, China, in 1999. He received the B.S. degree in automation from Lishui College, Lishui, China, in 2021 and the M.S. degree in control science and technology from the Zhejiang University of Technology, Hangzhou, China, in 2024.

He is currently an Algorithm Engineer with the Binjiang Institute, Zhejiang University, Hangzhou, China. His research interests include deep learning, machine learning, and their applications in different fields.



Jinyin Chen received the B.S. degree in computer science and the Ph.D. degree in control science and technology from the Zhejiang University of Technology, Hangzhou, China, in 2004 and 2009, respectively.

She studied evolutionary computing at the Ashikaga Institute of Technology, Ashikaga, Japan, in 2005 and 2006. She is currently a Professor with the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology. Her research

interests include artificial intelligence security, graph data mining, and evolutionary computing.



Jianwei Yin (Senior Member, IEEE) received the Ph.D. degree in computer science from Zhejiang University (ZJU), Hangzhou, China, in 2001.

He was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, GA, USA. He is currently a Full Professor with the College of Computer Science, ZJU. Up to now, he has authored or coauthored more than 100 papers in top international journals and conferences. His current research interests include service

computing, artificial intelligence, deep learning, and business process Management.

Dr. Yin is currently an Associate Editor for IEEE TRANSACTIONS ON SERVICES COMPUTING.