Non-Intrusive Load Monitoring with Low Carbon Technologies

Linshuo Zhang

Master of Science
School of Informatics
University of Edinburgh
2023

Abstract

With the increase of electric appliances usage in the home, the demand for electricity has surged. In response, the Non-Intrusive Load Monitoring (NILM) method was introduced to monitor individual appliances by analyzing a household's overall electricity consumption. In recent years, deep learning techniques, particularly the dilated Fully Convolutional Network, have been harnessed to address this task, exhibiting state-of-the-art disaggregation performance on the IDEAL dataset. Concurrently, with the rise of low-carbon technologies, various low-carbon appliances have become prevalent among households. However, the IDEAL dataset does not encompass data related to these low-carbon appliances. To bridge this gap, our project incorporated electric vehicle charging data into the original IDEAL dataset. Subsequent training and evaluation of the aforementioned FCN on this augmented dataset revealed that the FCN architecture retains its superior disaggregation performance for a majority of electric appliances. However, for appliances that possess similar power usage features to EVC, particularly those with consistent high power consumption like showers, the FCN's performance experiences a notable decline.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. The datasets used in this project are anonymized so that it did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Linshuo Zhang*)

Acknowledgements

I would like to express my gratitude for the help and time of my supervisor, Dr. Nigel Goddard. In addition, I am also deeply thankful to Dr. Elaine Farrow and Dr. Jonathan Kilgour, who were very helpful with the usage of models and datasets. Furthermore, I am indebted to the Sustin Lab for providing the model code and dataset. Finally, I would like to thank my family for their moral and material support during this time.

Table of Contents

Chapter 1

Introduction

The growing utilization of electricity applications has led to escalating demand for electricity[24], particularly in residential settings, where energy consumption currently constitutes between 20 and 40 percent of global energy usage[6]. As the number of electric appliances and devices within households increases, the management and efficient utilization of these resources demand greater attention and time. Nowadays, the primary means of obtaining statistics on electrical appliances consumption are monthly billing statements, which lack the granularity required to examine individual applications comprehensively. To address this issue and allocate appliance usage more accurately and efficiently, there is a pressing need for real-time power consumption monitoring. One naïve approach to achieve this is Intrusive Load Monitoring (ILM), which involves the installation of dedicated sensors for each appliance to detect its power usage. However, ILM proves costly due to equipment and installation expenses. As an alternative, Non-Intrusive Load Monitoring (NILM) has been proposed. NILM allows for the analysis of overall electricity consumption in a household, enabling the tracking of power usage for individual appliances without the need for dedicated sensors[24].

Furthermore, the phenomenon of population aging has resulted in a notable rise in the number of elderly individuals living independently[20]. This demographic shift has placed significant strain on socio-medical centers, which must now contend with an increased demand for elderly care and support. To address this challenge, the concept of remote supervision for the elderly has emerged as a potential solution. Smart meters equipped with Non-Intrusive Load Monitoring algorithms offer a promising avenue for achieving remote supervision. By leveraging the power consumption data collected by these smart meters, caregivers and healthcare professionals can gain valuable insights into the daily routines and activities of elderly individuals.[7] This real-time monitoring facilitates the detection of anomalies or irregularities in their power usage patterns, providing early indications of potential health issues or emergencies.

The progression of machine learning, specifically deep neural networks, has significantly facilitated the implementation of Non-Intrusive Load Monitoring (NILM) techniques[28, 4, 10]. In 2018, Brewitt and Goddard[2] introduced a cutting-edge NILM methodology based on Fully Convolutional Networks(FCN). This approach demonstrated exceptional effectiveness in accurately disaggregating the individual electricity consumption of appliances from the overall household power usage within the IDEAL dataset.

Carbon emission stands as a primary contributor to climate change and global warming. In a concerted effort to counteract these adverse effects and diminish reliance on fossil fuels, several low-carbon appliances have been developed. These innovations, including heat pumps, electric vehicles, and solar panels, are swiftly gaining traction in domestic settings. Given this backdrop, our project elected to focus on electric vehicle charging as a representative low-carbon appliance. The main objective was to ascertain whether the aforementioned FCN model retains its exemplary performance when the original IDEAL dataset is augmented with EVC data to simulate a household with an electric vehicle that is charged at home.

By employing the same FCN architecture and training and testing it on both the original and the augmented IDEAL datasets, we embarked on a comparative analysis of performance metrics among these configurations. Our findings indicated that apart from appliances like showers, which have similar power consumption patterns of EVC, the FCN framework demonstrated adaptability to the augmented dataset incorporating electric vehicle charging data.

Among the contributions we made in this project are:

- By integrating electric vehicle charging data provided by the Electric Nation into the raw IDEAL data, an augmented dataset of household electricity consumption containing EVC data was generated.

- The FCN models were trained, and their disaggregated performances were evaluated on the augmented dataset.

- Upon evaluating models trained on the original dataset with the augmented dataset, it was discerned that the capability of the original model, when evaluated on the original dataset, to disaggregate electric appliances was diminished.

- Upon contrasting the performance disparities between models trained on the original dataset and those trained on the augmented dataset, it was discerned that the FCN architecture retained robust disaggregation efficacy on the augmented dataset for the majority of electric appliances. Nonetheless, for appliances exhibiting sustained high electrical power consumption, the disaggregated capacity of the FCN architecture experienced a marked reduction.

Chapter 2

Background and Literature Review

2.1 Previous Works

The concept of Non-Intrusive Load Monitoring (NILM) was first introduced in 1985 by G. W. Hart[11], who proposed a methodology for identifying the operational state of household appliances by checking transient changes in power consumption. This approach hinged on the unique electrical ”signature” that each appliance shows upon activation. Hart’s seminal work laid the groundwork for subsequent research in the field of NILM. In recent years, the advent and development of machine learning have prompted the application of numerous deep learning models within the realm of NILM[28, 4, 10]. These advanced computational tools have expanded the capabilities of NILM, enabling more sophisticated and accurate analyses of appliance usage patterns and energy consumption.

Deep neural networks (DNNs) are available in a variety of structures and sizes tailored to their specific use cases[18]. They continue to evolve, improving the accuracy and efficiency of models. Representational networks possess the capability to render the inputs of DNN more abstract and meaningful through analysis and manipulation[27]. The nature of these inputs can vary widely, contingent upon the specific task, such as the pixels of an image[3, 13] and the amplitudes in an audio recording[1] and etc.

DNNs exhibit exemplary efficacy in processing unlabeled and unstructured data. Two preeminent architectures within the DNN framework are the Convolutional Neural Network (CNN)[15, 22, 12, 27] and the Recurrent Neural Network (RNN)[5, 23]. CNNs are typically delineated by three primary components: the convolutional layer, the pooling layer, and the fully connected layer. The convolutional layer which is the core of CNN, discerns intricate patterns within the data by executing convolutional operations on the input through one or more filters, culminating in a feature map. The pooling layer diminishes the sizes of these feature maps, retaining only key information. This not only reduces computational demands but also mitigates the potential for overfitting, thereby enhancing the network's efficiency. The fully connected layer is tasked with transposing these feature maps into the label space, allowing the model to produce corresponding outputs for different tasks[8]. Given the absence of a memory mechanism in this DNN, its current outputs usually don't depend on the sequence of previous inputs. To address this limitation, RNNs have been introduced, specifically tailored for sequential data processing. RNNs are characterized by an intrinsic recurrent architecture that amalgamates the current input with the preceding time step's hidden state, engendering a new hidden state. This mechanism of state propagation and recurrence structure equips the RNN with the capability to process sequential data, assimilating antecedent information and integrating it into current computations[23]. Such an architecture is adept at discerning temporal information and contextual interrelations within input sequences. Owing to these properties, both CNNs and RNNs have garnered significant attention and have found widespread applications in the fields of computer vision[12], natural language processing[16], and speech recognition [1].

Kelly and Knottenbelt (2015) employed multiple deep neural networks to execute the disaggregation of electricity consumption for various appliances, utilizing the UK-DALE dataset[14]. The aggregated dataset was synthesized as an input to the model through a sliding window technique, with the model subsequently predicting outputs that corresponded in size to the aforementioned input window. For baseline comparisons, two non-artificial intelligence methodologies, namely combinatorial optimization and the factorial hidden Markov model, were employed as well. Experimental findings indicated that, in contrast to these conventional methodologies, all DNN approaches achieved superior F1-score disaggregation performance. Notably, the most effective approach was a CNN-based denoising autoencoder. Such findings underscored the potential of DNNs as a promising avenue for future research in the realm of NILM tasks. However, a limitation was observed: the sliding window technique led to multiple processing of identical segments across different input sequences, resulting in computational redundancy and decelerated training. In a subsequent study, Brewitt and Goddard(2018) introduced a fully convolutional network scheme invoking dilated convolution[2]. This innovative approach circumvented the issue of overlapping outputs, thereby significantly enhancing training efficiency. Furthermore, their network's disaggregation performance was observed to surpass that of the Sequence to Point

Model[28], whose performance was at the forefront during that period. The current project employed this FCN model, and the following three sections provided a detailed description of its network structure and related techniques.

2.2 Dilated Convolutions

In conventional convolution operations, both the size and stride of the convolutional kernel remain constant, dictating the size of the receptive field and the resultant output feature map. However, the introduction of the dilation rate as an additional parameter in dilated convolution offers a novel approach. This parameter permits the insertion of gaps within the input feature map, thereby expanding the convolutional kernel's receptive field on the input feature map while maintaining the output feature map's size. As the dilation rate escalates, a filter of consistent size can encompass an expanded segment of the input sequence. Unlike pooling layers, dilated convolution does not lose information from the input, making it particularly advantageous for high-density prediction tasks, such as hyperspectral image classification[25] and speech emotion recognition[19], the dilated network can capture the contextual information in the input very effectively.

2.3 Fully Convolutional Networks

Fully Convolutional Networks (FCNs) were first introduced by Long and Shelhamer in 2015 [17]. In their seminal work, they elucidated the concept of FCNs and demonstrated its application to the domain of semantic image segmentation. Their methodology not only achieved state-of-the-art segmentation results but also accelerated training speed on the PASCAL VOC, NYUDv2 and SIFT Flow datasets. This pivotal contribution has significantly advanced the field of image segmentation, cementing the role of FCNs as an integral component within the realm of computer vision. The FCN, as a derivative of the conventional convolutional neural network, supplants the fully connected layers typically found in standard convolutional neural networks with convolutional layers. Consequently, the output manifests as a feature map, which, owing to its translation invariance, renders the employment of sliding windows in the FCN model computationally redundant.

2.4 Fully Convolutional Model

The structure of the FCN model network proposed by Brewitt and Goddard[2] was shown in Figure 2.1. All layers, with the exception of the final one, were configured with 128 filters. This quantity was manually selected as the maximum number of filters that would not significantly impede the training speed. The initial layer was a convolutional layer with a filter width of 9. The larger filter widths employed in this layer, relative to subsequent layers, served to balance the distribution of parameters across the layers. Following the initial layer, the network architecture incorporated a series of 9 dilated convolutional layers. Each of these layers was assigned a filter width of 3, the smallest filter size capable of accounting for past, present, and future information. The dilation rate for the first dilated convolutional layer was set to 2, with each subsequent layer featuring a dilation rate twice that of its predecessor. This configuration allowed for exponential growth of the network's receptive field while the number of layers increased linearly. Subsequent to the dilated convolutional layers, the network employed an additional convolutional layer with 128 filters and a width of one to further refine the output of the preceding layers. Finally, to condense the network output to a single channel, the last layer was configured with a single filter of width 1. This intricate architecture was designed to optimize the balance between computational efficiency and model performance.

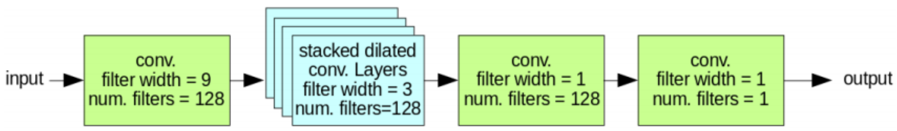

Figure 2.1: Fully convolutional network architecture[2]

2.5 Non-Intrusive Load Monitoring Task

NILM task is to deduce the energy consumption of individual household appliances and their distinct energy consumption stages over temporal intervals by analyzing the aggregated household electricity consumption.

2.6 Low Carbon Technology and Electric Vehicle Charging

”Low-carbon technology” refers to a broad spectrum of techniques aimed at minimizing carbon emissions and encouraging a shift to a low-carbon economy. Such technologies encompass a variety of methods, such as carbon-neutral and carbon-free alternatives, which significantly reduce greenhouse gas emissions, lessen the effects of climate change, and support sustainable development[26].

Electric vehicles are categorized as low-carbon appliances due to their reliance on electricity as the primary source of power. When compared with their conventional fuel-powered counterparts, electric vehicles are responsible for significantly lower carbon emissions, thereby contributing to a reduction in the overall environmental footprint. As the popularity of EVs continues to rise, an increasing number of drivers are opting for home-based charging solutions, caused by the considerations of cost efficiency and convenience[29]. This emerging trend underscores the escalating significance of domestic electric vehicle charging in household energy consumption. Hence, we selected EV charging as a representative of LC technology to evaluate the adaptability of the FCN-NILM model when the dataset was augmented with low-carbon technology data.

2.7 Dataset Description

2.7.1 Intelligent Domestic Energy Advice Loop Project

The Intelligent Domestic Energy Advice Loop (IDEAL) project embarked on an exhaustive data collection endeavor, amassing electricity and gas consumption data from a sample of 255 households within the United Kingdom. The temporal span of the data acquisition extended from a minimum of 55 days to a maximum of 673 days, exhibiting a mean duration of 286 days and a median of 267 days. It is of particular significance to note that individual appliance usage data was meticulously gathered from 39 of the participating residences, with measurements of electricity consumption documented at intervals of either 1 second or 5 seconds[21].

2.7.2 Electric Nation Project

The Electric Vehicle Charging dataset was derived from the Electric Nation Project, an initiative that amassed data from 673 EV chargers over the course of January 2017 to December 2018. Two commercial entities, CrowdCharge and GreenFlux, were responsible for monitoring the consumption of the EV charging, recording the data at resolutions of 1 minute and 3 minutes, respectively. Within the scope of this project, the data from CrowdCharge was selectively augmented into the IDEAL dataset. This decision was predicated on the higher resolution of CrowdCharge's data, which consequently enhances the model's ability to discern the characteristics of the EVC's power consumption more accurately. Furthermore, the utilization of Coordinated Universal Time (UTC) for timestamps in both the CrowdCharge recorded data and the IDEAL dataset facilitated the augmentation process, ensuring temporal consistency[9].

Chapter 3

Methodology

3.1 Data Preparation

3.1.1 Method for Augmenting IDEAL Dataset

Initially, the preprocessing of the EVC data was necessitated due to the temporal disparities in the installation of charging posts. To ensure the richness and comprehensiveness of the EVC data, the datasets with time intervals exceeding one year were selectively retained. This filtering process aimed to focus on the EVC data that provides a more extended and meaningful duration of charging activity, allowing for more appropriate integration into the IDEAL and more insightful analysis. Consequently, several charging data comprising 365 days of charging consumption for each were obtained, with each data point representing a time resolution of 1 minute. To align the data resolution with that of the IDEAL dataset (which maintains a resolution of 1 second), an upsampling technique was applied to the EVC dataset. This method allowed for the transformation of 1-minute resolution EVC data to 1-second resolution, ensuring compatibility with the IDEAL dataset. Finally, the year information was anonymized. By adopting this comprehensive approach, it was guaranteed that each timestamp in the IDEAL dataset corresponds uniquely to a counterpart in the EVC dataset, thus ensuring seamless integration and preserving the integrity of the dataset.

The data captured by sensors in the Electric Vehicle Charging (EVC) dataset delineates the amperage utilized during the charging process of electric vehicles. According to [9], a simple method for computing power was proposed. This method determines power consumption by multiplying a default voltage (240 Volt) with the current amperage. Such a methodology was employed in the present study. However, it's imperative to note that this approach may not yield precise absolute energy consumption values, primarily because it operates under the presumption of a unity power factor. An alternative and potentially superior method would involve the utilization of instantaneous apparent root mean square (RMS) power usage, which could offer a more accurate representation of absolute energy consumption. Nevertheless, this alternative was not adopted in the current project.

The primary objective of this project is to scrutinize the capability of the Non-Intrusive Load Monitoring Fully Convolutional Network (NILM-FCN) model to sustain its superior disaggregation performance when applied to datasets that incorporate EVC data. Consequently, we postulated the presence of an electric vehicle in each residential unit. For each home in the IDEAL dataset, data from a randomly selected charger was amalgamated into the household electricity consumption by aligning the parameters of month, day, minute, and second. Table 3.1 provided the structured outline of the process.

However, our method exhibited a limitation in that corresponding weeks are likely to differ across various years, even when the dates remain identical. For instance, the first of July in 2023 falls on a Saturday, whereas the same date in 2022 is a Friday. This temporal discrepancy may engender a divergence between the augmented dataset and the actual data, as individuals' electricity consumption behavior tends to vary between weekends and weekdays. Such variations could introduce biases or inaccuracies in the analysis.

Table 3.1: Steps to augment the dataset

Step	**Description**
1	Filtering out charging data from the original EVC dataset with a duration exceeding one year.
2	Restricting the obtained data to a duration of one year, ensuring a consistent time frame.
3	Upsampling the EVC data to a 1-second resolution to align it with the IDEAL dataset's resolution.
4	Integrating the EVC data into the IDEAL dataset by matching timestamps based on month, day, hour, and minute.

3.1.2 Splitting Train and Test sets

The homes selected in the study referenced in [2] were partitioned into two distinct subsets: 18 homes were allocated to the training set and 8 to the test set. These 26 homes were characterized by the presence of at least one of the specified appliances and a data collection period exceeding 30 days. Table 3.2 shows the division of the dataset.

Indeed, the k-fold cross-validation methodology is deemed superior. This technique partitions the original dataset into k subsets, subsequently amalgamating them into diverse training and validation sets based on a specified ratio for multiple training and evaluation models. Through this methodology, potential discrepancies arising from suboptimal selection of training and validation sets are mitigated. Consequently, the k-fold cross-validation offers a more precise assessment of performance. For this project, we adhered to the aforementioned division configuration to maintain consistency with the literature[2].

For each specific electrical appliance, the corresponding number of households possessing such appliance was systematically recorded and presented in Table 3.3. In the augmented dataset, electric vehicle charger was incorporated as an additional appliance category.

Table 3.2: Home IDs for training and test sets

Dataset	Home IDs
Training	62, 65, 96, 105, 106, 128, 136, 145, 162, 168, 169, 175, 228, 231, 238, 255, 263, 328
Test	73, 171, 212, 227, 242, 249, 264, 266

3.1.3 Data Preprocessing

To enable the training of the FCN-NILM model, the raw data need to undergo a process of cleaning and resampling, following the approach delineated in [2]. The IDEAL dataset captured instantaneous apparent root mean square (RMS) power usage for the entire household at a 1-second resolution, utilizing two distinct sensors: a 30A sensor and a 100A sensor. When the power consumption was below 4 kW, the reading from the 30A sensor was employed; otherwise, the reading from the 100A sensor was used. Gaps in the readings might occur due to signal propagation issues or system downtime, and these were addressed through three distinct strategies:

Table 3.3: The counts of homes for each appliance category for the training and test sets.

Appliance Type	Training	Test	Total
Washing Machine	13	6	19
Microwave	12	5	17
Dishwasher	10	4	14
Electric Kettle	8	4	12
Electric Shower	6	3	9
Electric Cooker	5	2	7
Electric Vehicle Charger	18	8	26

- Substitution with Alternate Sensor:

  If a reading from the 30A sensor is missing and a corresponding reading from the 100A sensor was available for the same timestamp, the latter reading was used to fill the gap.

- Short Interval Gap Filling:

  For remaining gaps where the time interval of the missing data was less than 1 minute, the previous valid readings were employed to fill the gap.

- Long Interval Gap Ignorance:

  For gaps that spanned longer time intervals, they were disregarded during the testing and training phases of the model.

Electric appliance readings were measured by individual appliance monitors (IAMs) with a resolution of 1 second or 5 seconds. To minimize the data volume, readings were taken from the IAMs only when there was a change in appliance power usage or, at a minimum, once per hour if there was no change in power usage. Consequently, the gaps in the readings lasting for an hour or less were interpreted as periods during which the power consumption of the appliance remained constant, and these gaps were forward filled. Any gaps lasting longer than one hour were treated as missing data and were disregarded during the training and testing of the models.

Finally, prior to the initiation of model training, all sensor readings were conducted to a downsampling process. Each data point was resampled at 10-second intervals, with the resultant value representing the mean of the readings recorded within those ten seconds. This preprocessing step served to reduce the data's granularity while preserving its essential characteristics, thereby facilitating more efficient computation.

3.2 Experimental Design

To effectively evaluate the different performances of the models, the experiments were divided into three distinct groups. All sets of experiments employed the same FCN-NILM model architecture illustrated in Chapter 2, but they differed in terms of the datasets used for training and testing. Firstly, the models trained on the original IDEAL dataset and tested on the original IDEAL dataset served as the baseline. The purpose of building this baseline was twofold. One was that it ensured that the FCN-NILM model demonstrated robust disaggregation performance on the original IDEAL dataset, and the other was that it provided a comparative benchmark against which the performance of the new models could be evaluated. Secondly, the models were trained on the original dataset and tested on the augmented dataset, which incorporated Electric Vehicle Charging data. The purpose of this set of experiments was to simulate a situation in life. When a family that previously did not have an electric vehicle buys one, is it necessary to update the model of the original electrical appliances? Thirdly, the models were trained on the augmented dataset and tested on the augmented dataset. This experimental suite constituted the primary focus of this project's request. Since these experiments provided the performance of the FCN-NILM model architecture when trained on the augmented dataset. By comparing the first and third groups, we could discern the disparities in the performance of the FCN-NILM architecture when trained on the original and augmented datasets.

For the assessment of models trained on the original IDEAL dataset, we trained distinct models for each of the six electrical appliances: microwaves, electric showers, electric cookers, dishwashers, kettles, and washing machines. In the context of the augmented dataset, beyond these six electrical appliances, we also trained models specifically tailored for electric vehicle consumption patterns.

3.2.1 Experimental Setup

3.2.1.1 Hardware and Software

Four Nvidia Titan X GPUs were used to train the models in two groups of experiments. The processor was a CPU Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz. The operating system on which the program ran was Ubuntu 20.04.6 LTS. The deep learning library used was Tensorflow-gpu v2.4.1.

3.2.1.2 Hyper-Parameter Setting

ADAM optimizer was employed for training NILM-FCN models. The learning rate was initially set to 1×10^{-3}. This rate was subsequently reduced by a factor of 10 whenever the loss did not improve over a span of 10 epochs. The training process was terminated under two conditions: either when the validation error rate had not shown improvement for 15 consecutive epochs, or after the completion of a total of 200 epochs. Furthermore, a batch size of 256 was consistently used throughout the training process. This configuration of hyperparameters was chosen to optimize the balance between computational efficiency and model accuracy.

3.2.2 Evaluation Metrics

To assess the performance of the model within this project, three distinct metrics were employed: Mean Absolute Error (MAE), Signal Aggregate Error (SAE), and the F1 score. Both MAE and SAE are conventional evaluation metrics[2] utilized for Non-Intrusive Load Monitoring (NILM) tasks, concentrating on the discrepancies between the predicted values(i.e., power consumption) and the ground truth. Mathematically, MAE is defined as the average of the absolute differences between the predicted and actual values, while SAE may involve a more complex aggregation of errors. Conversely, the F1 score serves a different evaluative purpose, focusing on the classification aspect of the problem. Specifically, it assesses the model's ability to accurately determine whether the electrical appliances are in an active state or not. By employing these three metrics in tandem, a comprehensive and multifaceted evaluation of the model's performance can be achieved, encompassing both regression accuracy and classification precision.

3.2.2.1 Mean Absolute Error(MAE)

The mean absolute error (MAE) symbolizes the discrepancy between the actual value and the predicted value of appliance consumption at each time instance. It can be computed using Equation 3.1:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - x_t| \tag{3.1}$$

In this equation, x_t refers to the real consumption of an electrical appliance, while y_t represents its predicted consumption at time t[2].

3.2.2.2 Normalized Signal Aggregate Error(SAE)

Along with tracking the electrical consumption at individual moments in time, it's vital to account for the error in the cumulative electrical consumption over a specific duration. The overall precision of the consumption estimate might be acceptable, even if there are inaccuracies in the predictions at particular time steps. The metric for this can be expressed using Equation 3.2:

$$SAE = \frac{|\sum_t y_t - \sum_t x_t|}{\sum_t x_t} \tag{3.2}$$

Here, $\sum_t x_t$ signifies the total of the true values of the electrical appliance's consumption within a defined period, while $\sum_t y_t$ denotes the total of the predicted values for that same interval[2].

3.2.2.3 F1-Score

The F1 score is a widely recognized metric for evaluating the performance of classification models, particularly in scenarios where the distribution of classes is imbalanced. It is a harmonic mean of precision and recall, providing a balanced model accuracy assessment. The mathematical expression for the F1 score is given as Equation 3.3

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.3}$$

In the context of a binary classification problem, the four possible classifications of prediction results are:

- True Positive (TP): The instance is positive, and the prediction is correct.
- False Positive (FP): The instance is negative, and the prediction is incorrect.
- True Negative (TN): The instance is negative, and the prediction is correct.
- False Negative (FN): The instance is positive, and the prediction is incorrect.

Precision is the ratio of the number of true positive predictions, and the number of all predicted positive samples. It measures how many positive class predictions are correct. The formula for the calculation can be represented as Equation 3.4

$$Precision = \frac{TP}{TP + FP} \tag{3.4}$$

Recall is a metric that evaluates the ability of a classification model to identify the positive class instances correctly. It is defined as the ratio of the number of true positive predictions and the number of actual positive samples, representing the proportion of actual positive class instances that the model correctly predicted. Mathematically, recall is expressed as Equation 3.5:

$$Recall = \frac{TP}{TP + FN} \tag{3.5}$$

Within the scope of our project, household appliances are classified as the positive class when activated and the negative class when deactivated. By establishing specific rules for different appliances, the operational state of an appliance could be inferred from the model's output. These rules delineated the maximum and minimum durations for which an appliance is active, the interval between two consecutive activations, and the minimum power consumption required to ascertain that an appliance is active.

3.2.3 Setting of Appliance Rules

Table 3.4 provided descriptions of the rule parameters, while Table 3.5 enumerated the predefined values for each appliance rule. A critical aspect of this project involved the determination of appropriate rules for electric vehicle charging.

For the parameter setting of the EV charging, the value will depend on a variety of factors, including the battery capacity of the EV, the maximum power of the charger, the capacity of the power grid, and the user's demand. With the information from [9], we had the following settings for the EVC rules:

- **min_on_duration:** 600 seconds

  In everyday scenarios, the power of an electric vehicle charger may experience transient fluctuations due to regular variations in the electrical grid or interference from other electronic devices. Such fluctuations could erroneously lead to the assessment that the charger is in an active state. To mitigate this potential source of error, a minimum activated time threshold can be established. By setting this min_on_duration to 10 minutes, brief and inconsequential fluctuations can be effectively disregarded, enhancing the accuracy of the charger's operational state determination.

- **min_off_duration:** 900 seconds

  Analogous to the parameter max_off_duration, the setting of min_off_duration prevents a brief reduction in power from leading to an incorrect assessment that the charger has transitioned to an inactive operational state. This consideration is vital in accurately capturing the true behavior of the charging process, particularly in the presence of transient power fluctuations. Furthermore, to protect the electric vehicle battery, it is essential to allocate a cooling-down period prior to engaging in the next charging behavior. By stipulating 15 minutes to min_off_duration, the model can more accurately distinguish whether a given scenario represents one fluctuating charge or two distinct consecutive charges.

- **max_on_duration:** 18000 seconds

  For the majority of electric vehicles, a charging duration of approximately 5 hours is typically sufficient to replenish the battery to its full capacity.

- **on_power_threshold:** 1500 Watts

  In the context of electric vehicle charging, a threshold value of 1.5 kW is set to ascertain whether the charger is activated. This value serves as a reasonable and broadly applicable benchmark, alignable with the capabilities of most contemporary EV chargers. If the power consumption of the charger falls below this threshold, it may suggest that the battery has reached its full charge capacity, or it could signal the presence of some underlying issues.

Table 3.4: Rule parameters descriptions

Rule Parameters	**Description**
min_on_duration	The minimum duration for which an appliance is active.
min_off_duration	The interval between two consecutive activations of the appliance.
max_on_duration	The maximum duration for which an appliance is active.
on_power_threshold	The minimum amount of power consumption of the appliance during an activation.

Table 3.5: Rule parameters setting for household appliances[2]

appliance	**min_on_duration**	**min_off_duration**	**max_on_duration**	**on_power_threshold**
microwave	10s	15s	10000s	200W
electric shower	20s	30s	5000s	300W
electric cooker	10s	120s	10000s	200W
dishwasher	600s	400s	15000s	50W
kettle	10s	10s	600s	1000W
washing machine	1200s	600s	30000s	50W
electric vehicle charger	600s	900s	18000s	1500W

Chapter 4

Results and Discussion

In this chapter, the model performance of the seven appliances was stated individually. Moreover, six appliances selected from the original dataset were analyzed for the performance difference between the models trained on the original dataset and the models trained on the augmented dataset. This comparison aimed to elucidate the ramifications of training the FCN model on the augmented dataset in terms of disaggregation efficacy for the original appliances. With respect to the EVC, the discourse explored the model's proficiency in disaggregating this specific low-carbon technology. In subsequent delineations and figures, the notations "ori_ori," "ori_aug," and "aug_aug" were employed to respectively signify the first, second, and third experimental sets as outlined in Section 3.2.

4.1 Dishwasher Prediction

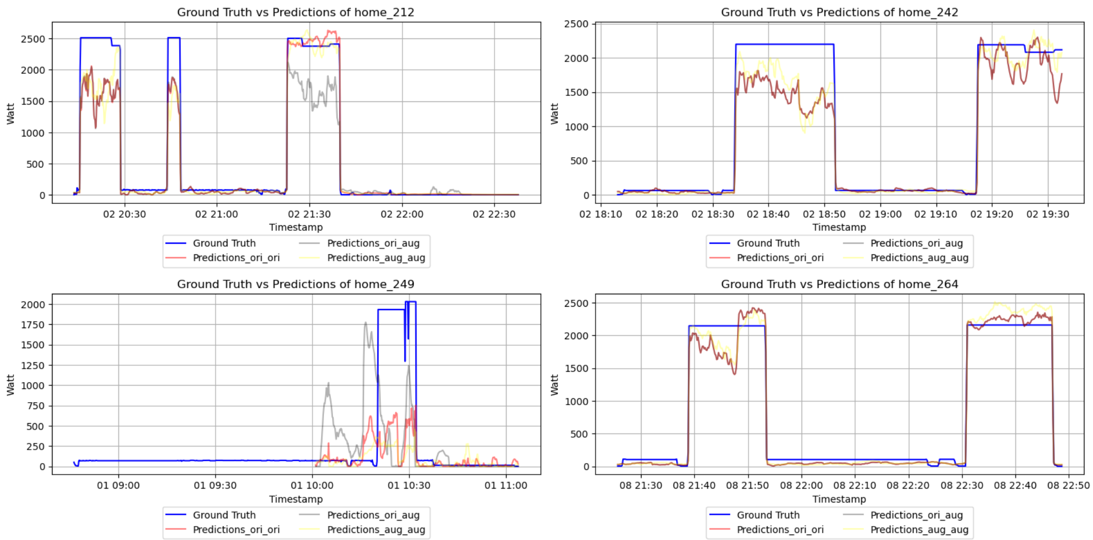

Figure 4.1: The predictions of dishwasher

For the dishwasher, each of the three models indicated activation in several instances as depicted in Figure 4.1. Concurrent with the peak of the ground truth curve, all three models exhibited a pronounced elevation. However, a discernible disparity existed between some of the predictions and the ground truth. As illustrated in Figure 4.2, the metrics of ”ori_ori” and ”aug_aug” were almost similar, while ”ori_aug” displayed a marginal decline in performance. Consequently, post the incorporation of EVC data, the FCN-NILM architecture retained its robust disaggregation capability for the dishwasher, and retraining the dishwasher model on the augmented dataset was imperative.

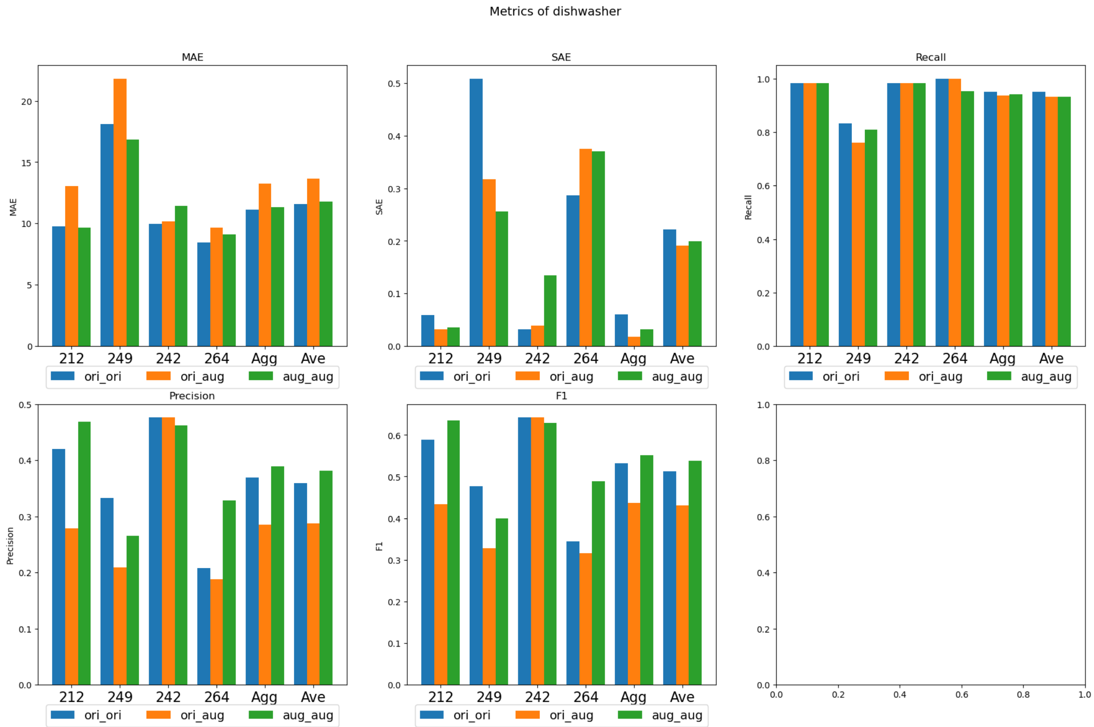

Figure 4.2: The metrics of dishwasher

4.2 Cooker Prediction

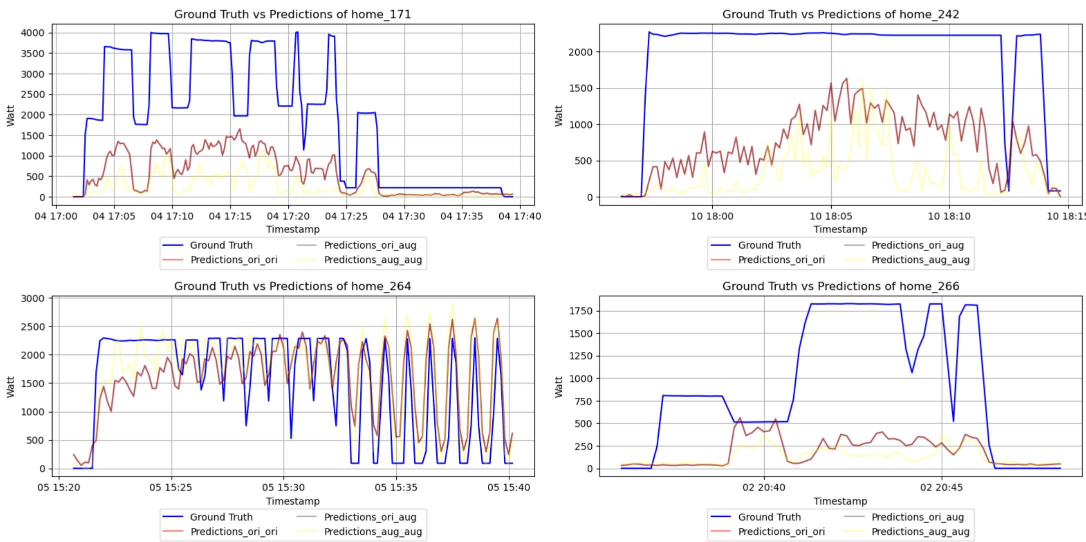

Figure 4.3: The predictions of cooker

Within the quartet of instances presented in Figure 4.3, all three models exhibited an upward bulge during the operation of the cooker. Nonetheless, several substantial gaps were observed between the predicted and actual values. Notably, the curves for ”ori_ori” and ”ori_aug” were perfectly overlapped during these instances, indicating that the integration of EV data had a negligible impact on the performance of the models trained on the original dataset. The evaluation metrics for the three models were delineated in Figure 4.4. Both SAE and MAE had large values, attributable to the pronounced deviation between the predicted and actual values. The cooker’s prediction was characterized by a high recall yet diminished precision, likely stemming from the cooker’s intricate power consumption pattern, influenced by its diverse operational states. This complexity might have led the model to overlook numerous transient states. Among the three models, the metrics remained relatively stable. Thus, the integration of EVC data exerted minimal influence on the cooker’s disaggregation performance for the models trained on the original dataset or the augmented dataset. Consequently, the FCN-NILM architecture remained viable for the cooker on the augmented dataset.

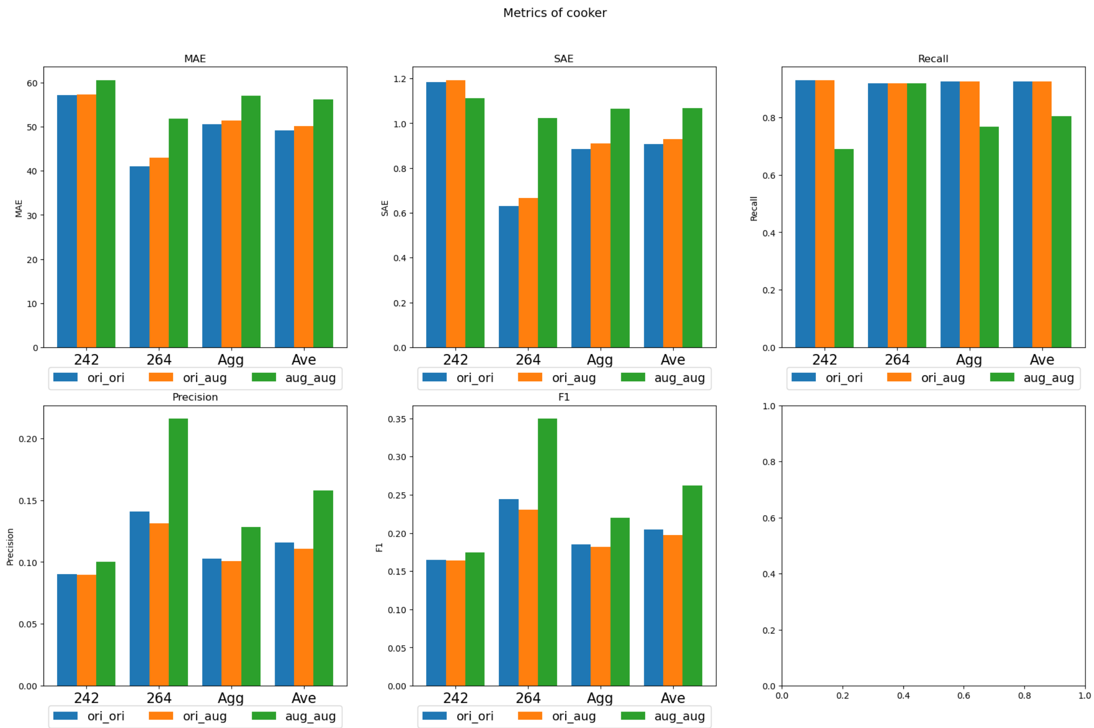

Figure 4.4: The metrics of cooker

4.3 Kettle Prediction

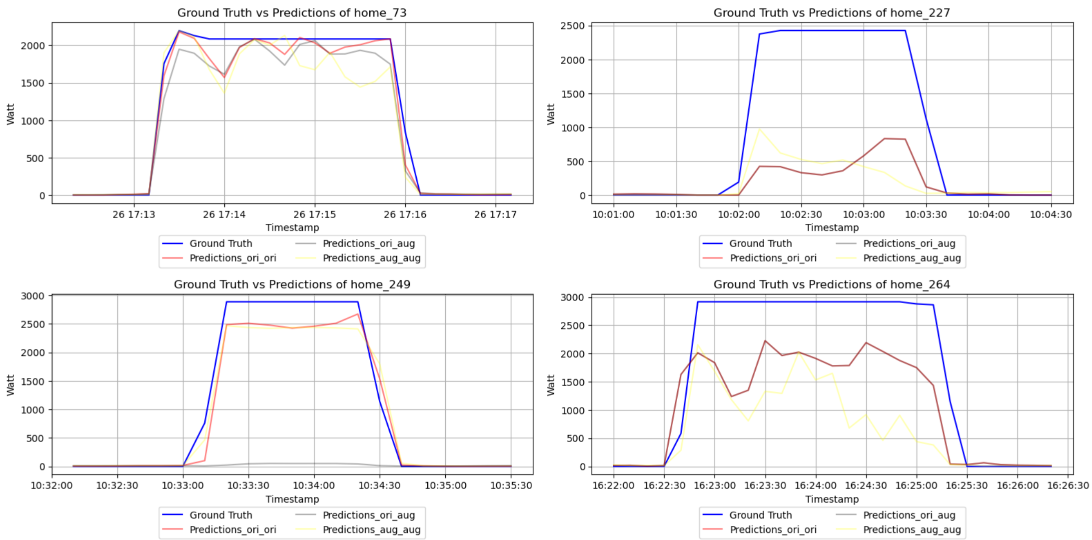

Figure 4.5: The predictions of kettle

In Figure 4.5, the predicted outcomes from both ori_ori and aug_aug demonstrated commendable accuracy across the three sampled homes. However, the ori_aug model showed a deficiency in accurately detecting the kettle's activation state in home 249. This discrepancy might arise from the kettle's distinctive features being overshadowed by the incorporation of the EVC data. As observed from the metrics presented in Figure 4.6, while ori_ori and aug_aug exhibited close similarities, the ori_aug indicated a marginal decline in the model's performance. Consequently, post the integration of EVC data, the FCN-NILM architecture retained its robust performance in disaggregating the kettle's consumption. What's more, it underscored the imperative of retraining the kettle model on the augmented dataset.

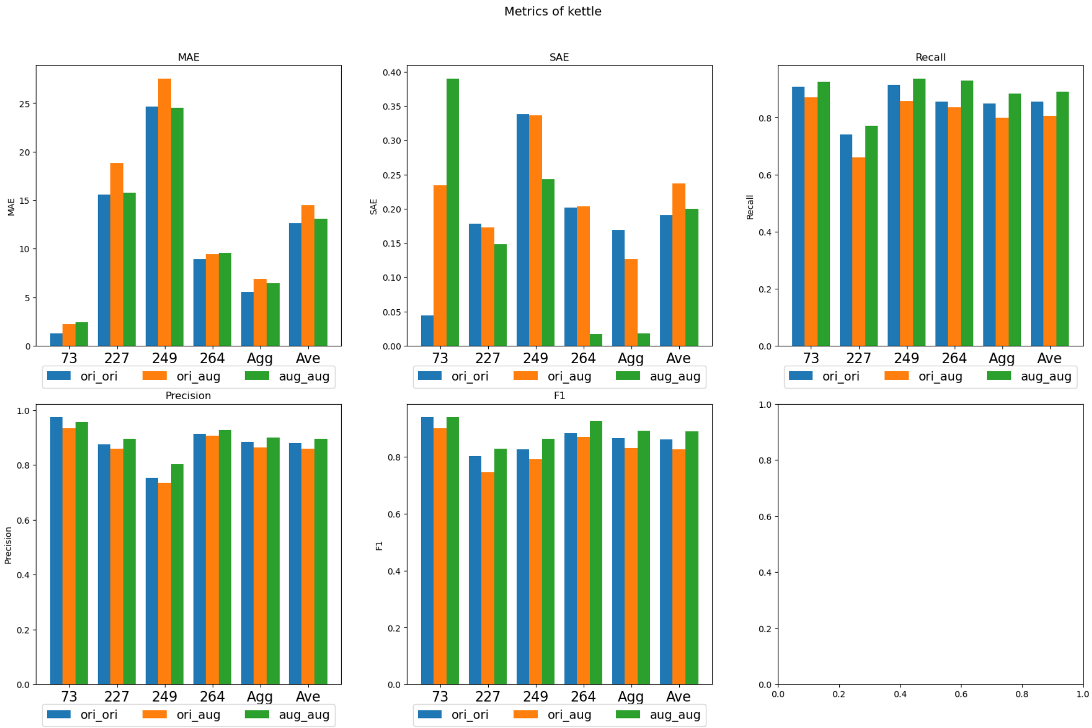

Figure 4.6: The metrics of kettle

4.4 Washing Machine Prediction

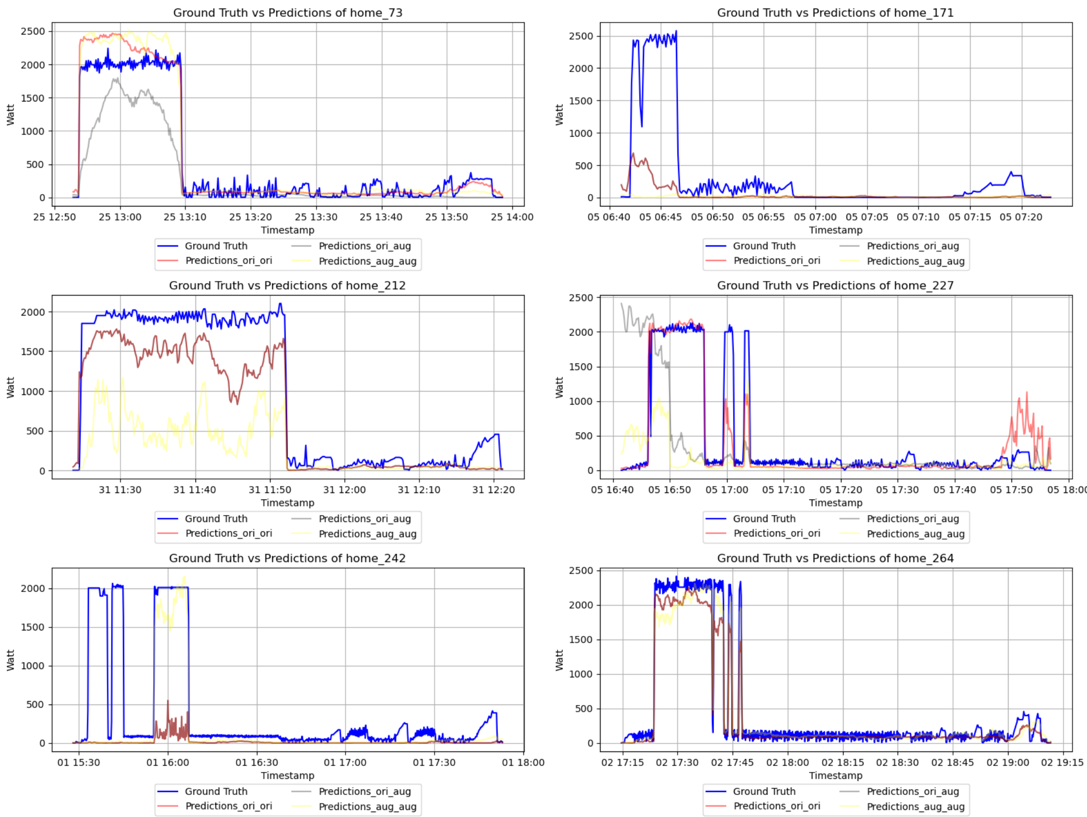

Figure 4.7: The predictions of washing machine

In Figure 4.7, all three models aptly predicted the activation state of the washing machine. However, gaps observed in certain instances between the predicted values and the actual ones. Analyzing the metrics presented in Figure 4.8 for the washing machine model, it was evident that the performances of ori_ori and aug_aug were closely aligned. In contrast, the ori_aug model introduced a minor decline in performance. Consequently, following the integration of the EVC data, the FCN-NILM architecture maintained its efficacy in disaggregating the washing machine's consumption. This observation underscored the importance of retraining the washing machine model using the augmented dataset.

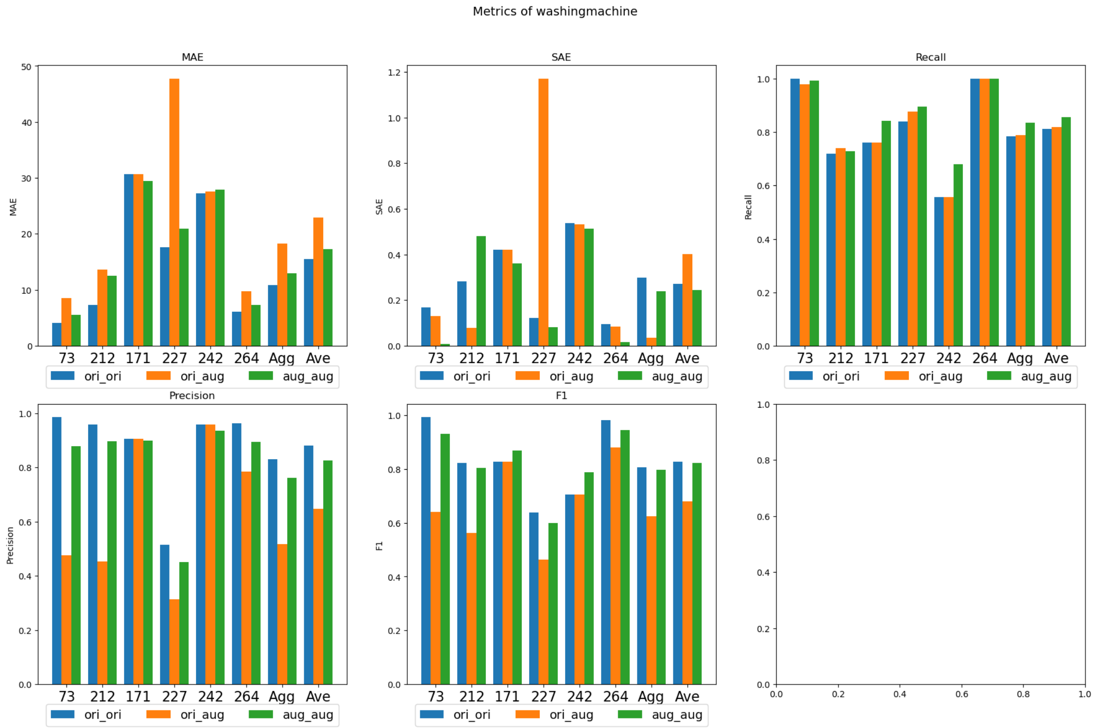

Figure 4.8: The metrics of washing machine

4.5 Shower Prediction

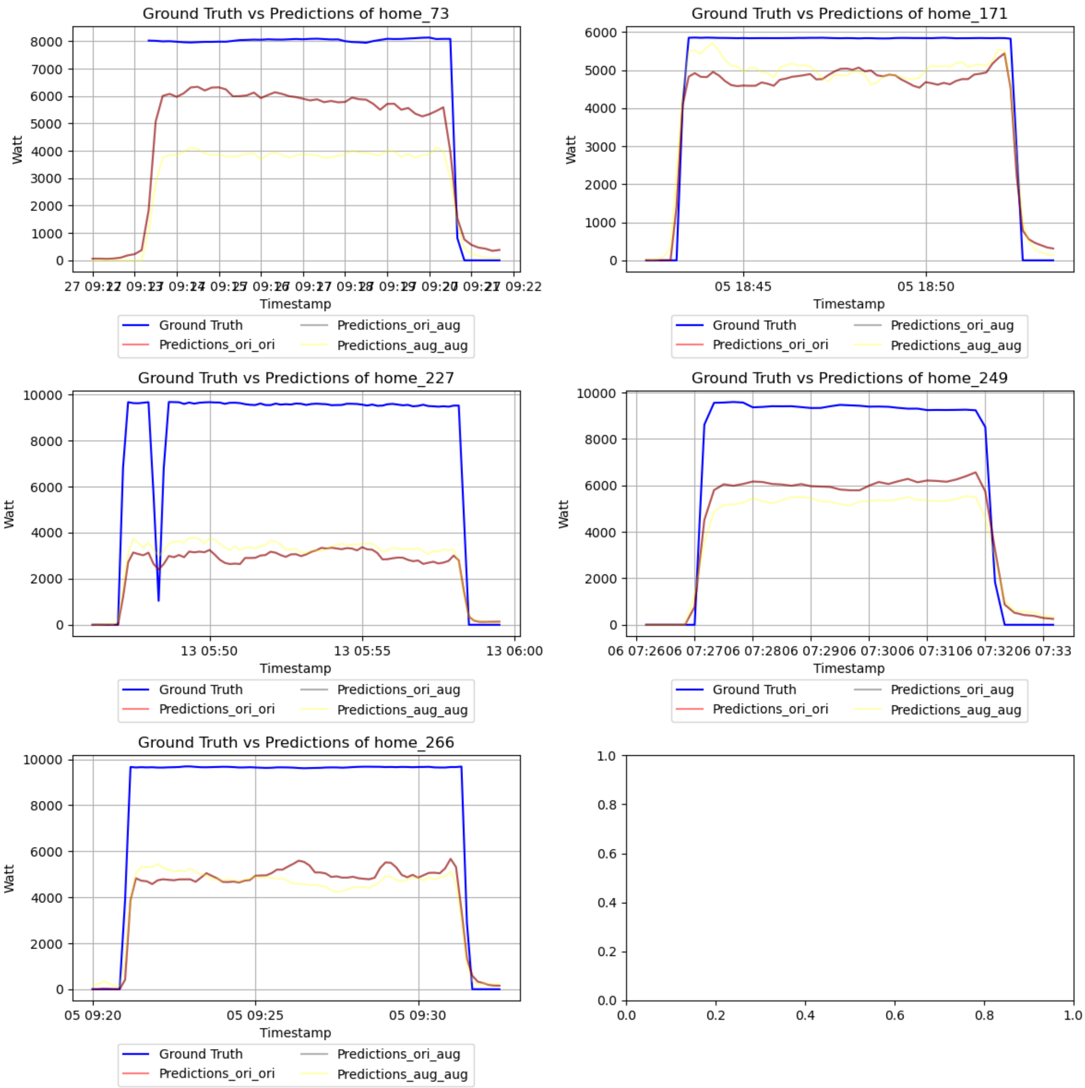

Figure 4.9: The predictions of shower

In Figure 4.9, during the active periods of the shower, all three models exhibited an upward trend. However, notable gaps were observed between the predicted and actual values. However, the discrepancies among the predicted values across the three models were relatively minor. Analyzing the metrics presented in the subsequent Figure 4.10 for the shower model, the ori_ori model demonstrated superior recall, precision, and F1 score. In contrast, models evaluated on the augmented dataset(i.e. ori_aug and aug_aug) manifested a pronounced decline in both precision and F1 score metrics. This degradation in performance might be attributed to the similar high power consumption characteristics of both the shower and EVC. Such property could potentially diminish the model's capacity to discern the unique features of the shower, leading to misclassification of non-shower patterns as shower-related. Consequently, the integration of EVC data appeared to significantly influence the disaggregation performance of the shower model. This suggested a potential need for optimization strategies tailored for the FCN-NILM architecture when applied to the shower.

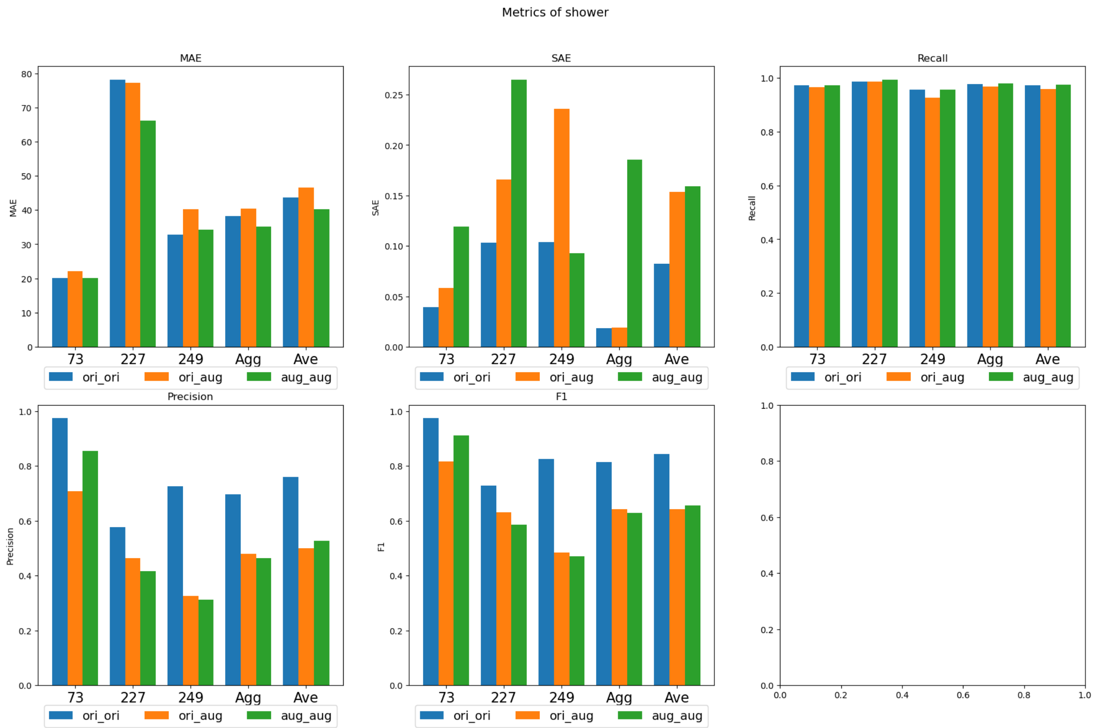

Figure 4.10: The metrics of shower

4.6 Microwave Prediction

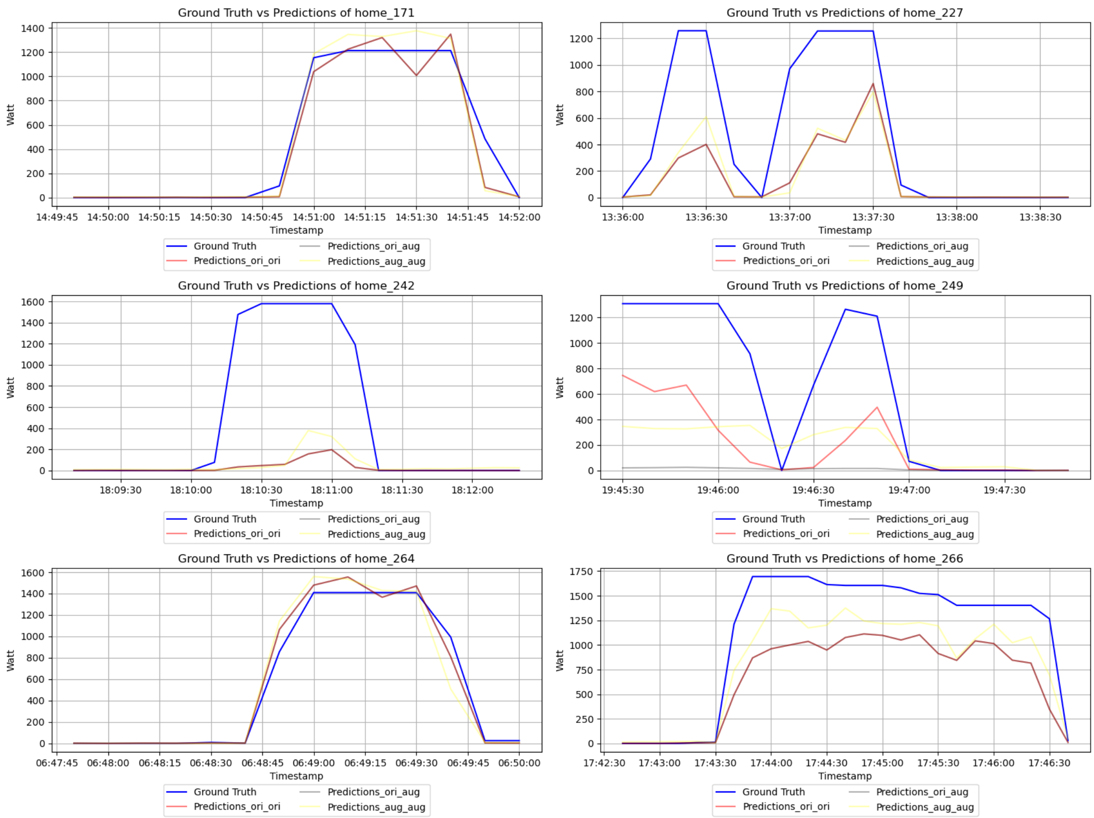

Figure 4.11: The predictions of microwave

In Figure 4.11, the three models aptly forecasted the activation state of the microwave. Nonetheless, certain instances revealed the gaps between the model's predictions and the ground truth. Analyzing the metrics delineated in Figure 4.12 for the microwave models, one could discern that despite the existence of some differences in the metrics among the three models, these deviations were inconsequential. Consequently, post the integration of the EVC data, the FCN-NILM architecture persevered in upholding its intrinsic efficacy in the disaggregation task of the microwave.

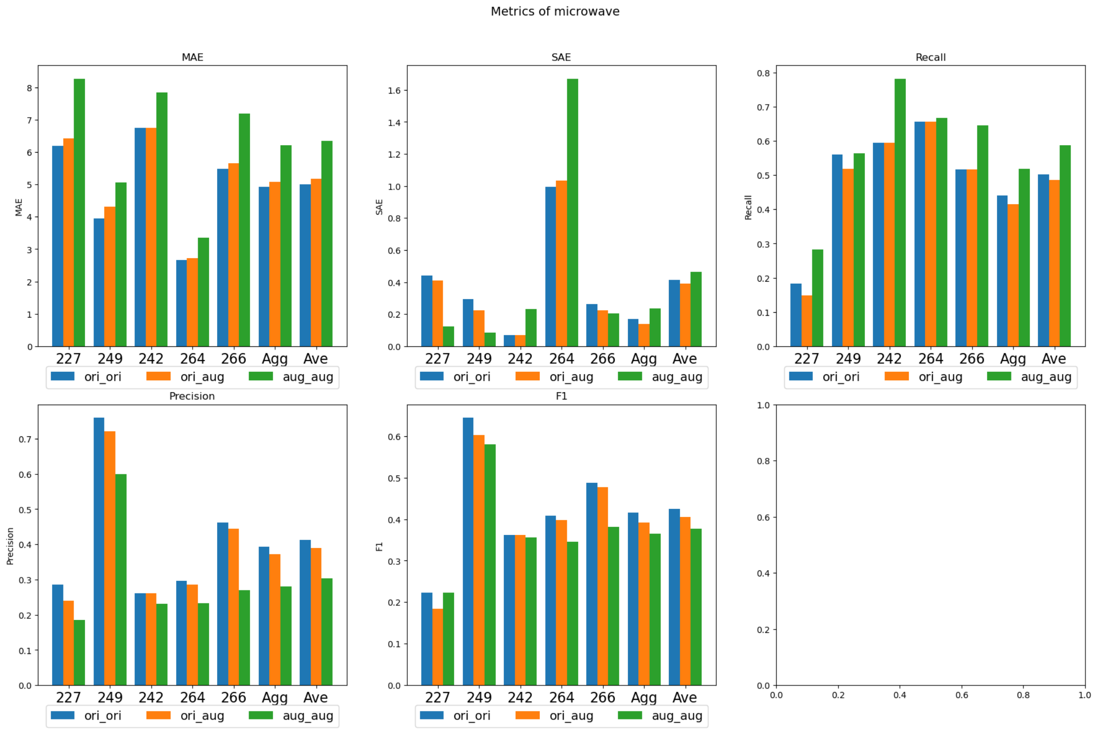

Figure 4.12: The metrics of microwave

4.7 Electric Vehicle Charging Prediction

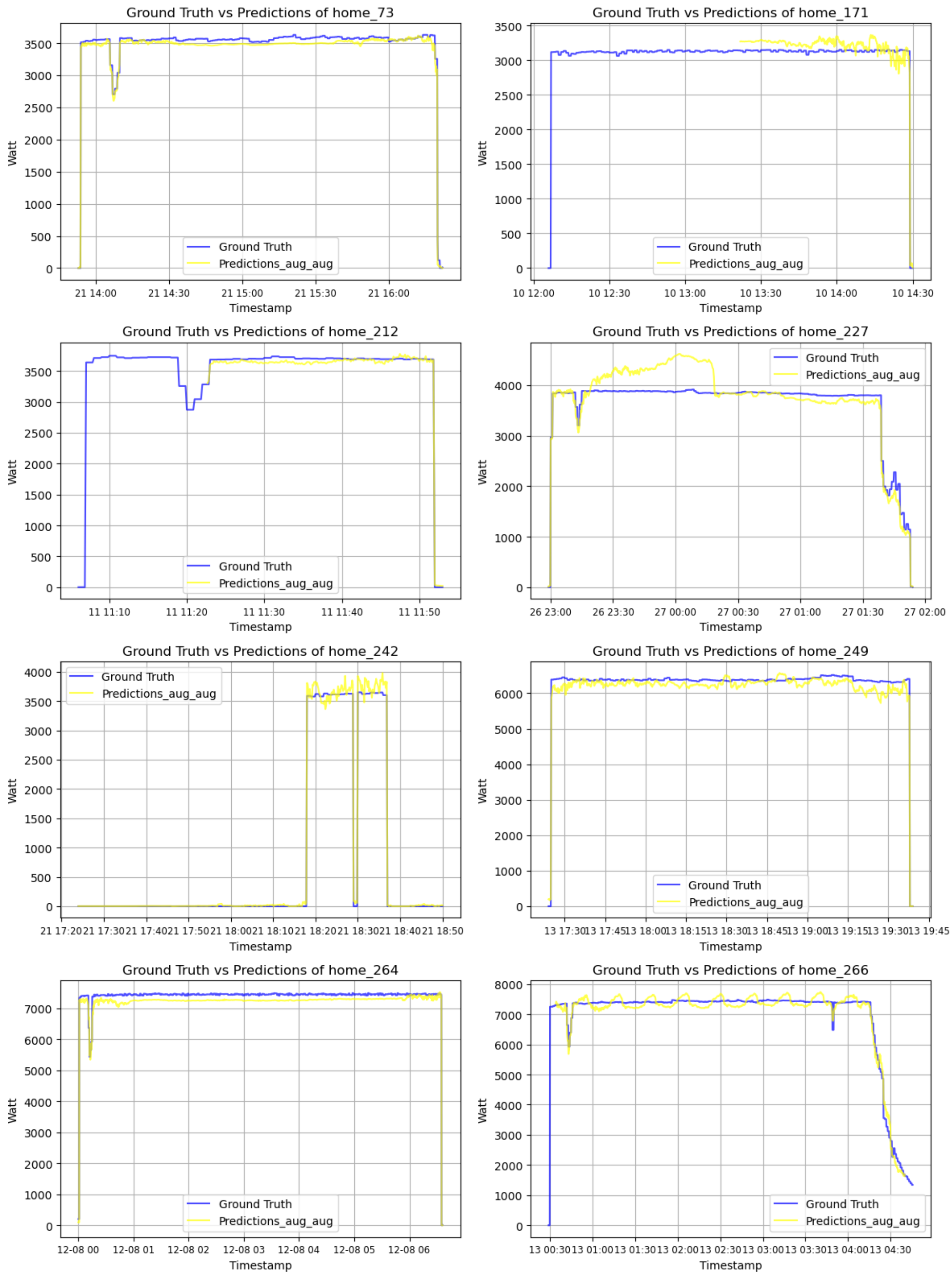

Figure 4.13: The predictions of electric vehicle charging

In Figure 4.13, the FCN-NILM model's predictions for electric vehicle charging closely aligned with the actual values. The metrics presented in Figure 4.14 for the electric vehicle charging model exhibited the exceptional performance of the FCN-NILM architecture in the disaggregation task for this specific application. The prolonged duration and high power consumption associated with electric vehicle charging, in contrast to other household appliances, likely rendered its pattern more distinguishable and thus more readily learned by the model. Consequently, the FCN-NILM architecture adeptly handled the disaggregation task for electric vehicle charging, a representative of low-carbon electrical appliances.

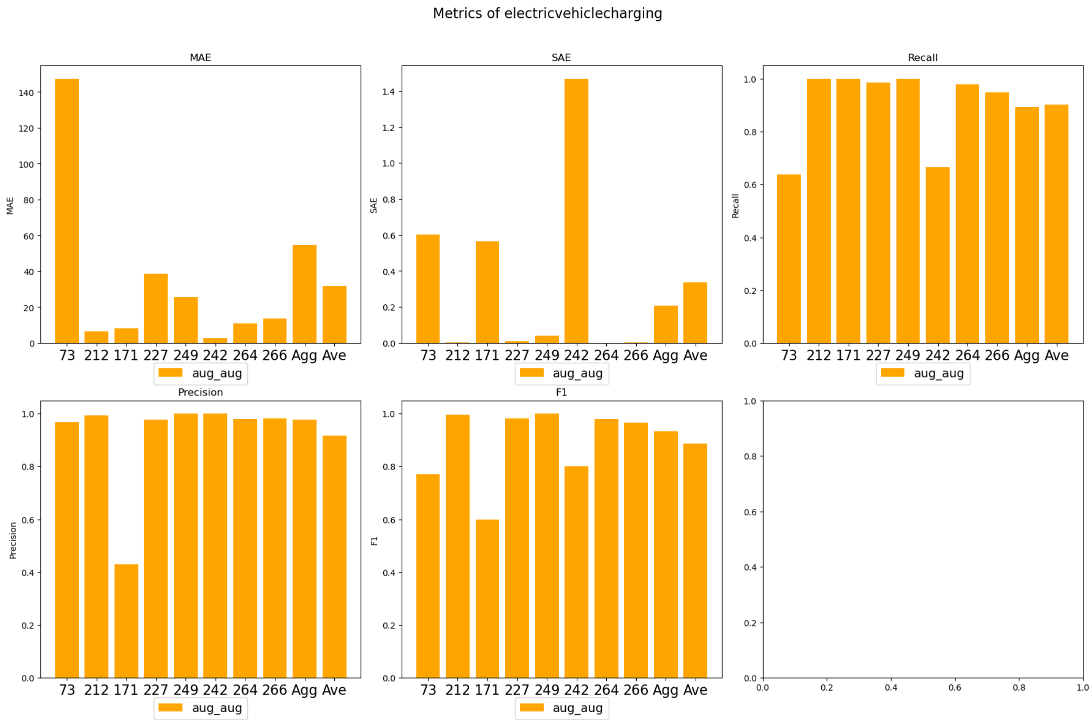

Figure 4.14: The metrics of electric vehicle charging

4.8 Discussion

In Table 4.1, the MAE and SAE metrics were presented, aggregated across all households for each appliance. Table 4.2 delineated the Recall, Precision, and F1-score metrics, also aggregated across all households for each respective appliance. The optimal evaluation results for each metric and appliance were emphasized in bold within these tables. Given that the primary objective of the model was to monitor the activation status of electrical appliances, the metrics in Table 4.2 had greater significance.

Upon analysis, the models trained on the augmented dataset incorporating EVC data exhibited marginal variations in their disaggregation capabilities for the five non-low-carbon appliances, with the exception of the shower. Specifically, for the shower, there was a substantial reduction in accuracy by 38.5% compared to the original, leading to a 22% decrease in the F1-score. This deviation might have been attributed to the analogous high power consumption patterns possessed by both the EVC and the shower. Furthermore, the performance metrics of the aug_aug models predominantly surpassed those of the ori_aug models, indicating the necessity for model updates whenever electric vehicles as a new appliance were introduced into a household.

Remarkably, the EVC achieved a recall of 89%, a precision of 98%, and an F1-score of 0.93 when trained and tested on the augmented dataset. This underscored the FCN-NILM's robust performance in disaggregating EVC power consumption from the overall power usage. In conclusion, barring the shower, the FCN-NILM architecture demonstrated commendable adaptability when trained and tested on augmented datasets that incorporated EVC data.

Appliances	MAE(watt)			SAE		
	ori_ori	ori_aug	aug_aug	ori_ori	ori_aug	aug_aug
Dishwasher	**11.11**	13.28	11.34	0.06	**0.02**	0.03
Cooker	**50.56**	56.98	51.46	**0.89**	1.06	0.91
Kettle	**5.56**	6.87	6.42	0.17	0.13	**0.02**
Washing Machine	**10.83**	18.22	12.99	0.30	**0.03**	0.24
Shower	38.17	40.35	**35.19**	**0.02**	**0.02**	0.19
Microwave	**4.92**	5.08	6.21	0.17	**0.14**	0.24
Electric Vehicle Charging	\	\	54.67	\	\	0.21

Table 4.1: Comparison of Original and Augmented Models for MAE and SAE

Appliances	Recall			Precision			F1		
	ori_ori	ori_aug	aug_aug	ori_ori	ori_aug	aug_aug	ori_ori	ori_aug	aug_aug
Dishwasher	**0.95**	0.94	0.94	0.37	0.29	**0.39**	0.53	0.44	**0.55**
Cooker	**0.93**	**0.93**	0.77	0.10	**0.13**	0.10	0.19	0.18	**0.22**
Kettle	0.85	0.80	**0.88**	0.88	0.86	**0.90**	0.87	0.83	**0.89**
Washing Machine	0.78	0.79	**0.84**	**0.83**	**0.83**	0.76	**0.81**	**0.81**	0.80
Shower	**0.98**	0.97	**0.98**	**0.70**	0.48	0.46	**0.81**	0.64	0.63
Microwave	0.44	0.41	**0.52**	**0.39**	0.37	0.28	**0.42**	0.39	0.36
Electric Vehicle Charging	\	\	0.89	\	\	0.98	\	\	0.93

Table 4.2: Comparison of Original and Augmented Models for Recall, Precision, and F1

Chapter 5

Conclusions and Future Work

5.1 Conlusions

In this research, the extant IDEAL dataset was enriched by incorporating electric vehicle charging data, as sourced from Electric Nation. Subsequent to this augmentation, the dilated Fully Convolutional Network (FCN) architecture was employed to train and evaluate models for six prevalent electric appliances including dishwasher, cooker, kettle, washing machine, shower and microwave, and a new low-carbon appliance, electric vehicle charging. The efficacy of these models was determined through a quintet of metrics: MAE, SAE, Recall, Precision, and F1. The salient outcomes of this investigation were delineated in the following:

- By comparing the performance of the models trained on the original dataset and tested on the original dataset with the models trained on the original dataset and tested on the augmented dataset, empirical findings elucidated that to preserve the robustness of the FCN architecture in executing the NILM task when a residence devoid of an electric vehicle subsequently procures one, it becomes imperative to recalibrate the antecedent models. Specifically, this necessitates the substitution of the models, which were trained on the original dataset, with their counterparts trained on the augmented dataset.

- By comparing the performance of the models trained on the original training set and tested on the original dataset with that of the models trained on the augmented dataset and tested on the augmented dataset, our observations revealed that barring the appliance 'shower', the FCN-NILM architecture consistently upholds its strong disaggregation capability for the remaining five electrical appliances when the augmented dataset is employed for training.

- Finally, the FCN-NILM architecture exhibited exemplary efficacy in the disaggregation task specific to electric vehicle charging, achieving 89% recall, 98% precision, and an F1-score of 0.93.

Thus, employing the dilated FCN scheme for the NILM task of most appliances on augmented household electricity consumption datasets, inclusive of EVC data, is indeed viable. However, it has to exclude electric appliances that manifest energy consumption patterns analogous to EVC, such as showers.

5.2 Future Work

In this study, we exclusively utilized electric vehicle charging as a representative of low-carbon appliances to investigate the efficacy of the FCN framework in addressing NILM tasks on datasets inclusive of low-carbon data. Consequently, assessing the performance of FCN models in handling NILM tasks on datasets that encompass data of multiple low-carbon appliances emerges as a compelling avenue for future research. Contemporary low-carbon appliances such as heat pumps and solar panels are gaining traction. Notably, solar panels present a unique challenge; unlike conventional electrical appliances, they contribute to a reduction in overall household electricity consumption due to their electricity generation capability. Moreover, the performance of the model for the shower when trained on the augmented dataset suffered from a significant decline. Therefore, determining strategies to mitigate this impact is another valuable orientation for future investigation.

Bibliography

[1] Dario Amodei, Rishita Anubhai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Jingdong Chen, Mike Chrzanowski, Adam Coates, Greg Diamos, Erich Elsen, Jesse Engel, Linxi Fan, Christopher Fougner, Tony Han, Awni Hannun, Billy Jun, Patrick LeGresley, Libby Lin, Sharan Narang, Andrew Ng, Sherjil Ozair, Ryan Prenger, Jonathan Raiman, Sanjeev Satheesh, David Seetapun, Shubho Sengupta, Yi Wang, Zhiqian Wang, Chong Wang, Bo Xiao, Dani Yogatama, Jun Zhan, and Zhenyao Zhu. Deep speech 2: End-to-end speech recognition in english and mandarin, 2015.

[2] Cillian Brewitt and Nigel Goddard. Non-intrusive load monitoring with fully convolutional networks, 2018.

[3] Jinyin Chen, Haibin Zheng, Hui Xiong, Ruoxi Chen, Tianyu Du, Zhen Hong, and Shouling Ji. Finefool: A novel dnn object contour attack on image recognition based on the attention perturbation adversarial technique. *Computers & Security*, 104:102220, 2021.

[4] Kunjin Chen, Qin Wang, Ziyu He, Kunlong Chen, Jun Hu, and Jinliang He. Convolutional sequence to sequence non-intrusive load monitoring, 2018.

[5] Weiling Chen, Chai Kiat Yeo, Chiew Tong Lau, and Bu Sung Lee. Leveraging social media news to predict stock index movement using rnn-boost. *Data & Knowledge Engineering*, 118:14–24, 2018.

[6] K.J. Chua, S.K. Chou, W.M. Yang, and J. Yan. Achieving better energy-efficient air conditioning – a review of technologies and strategies. *Applied Energy*, 104:87–104, 2013.

[7] Shuang Dai, Qian Wang, and Fanlin Meng. A telehealth framework for dementia care: an adls patterns recognition model for patients based on nilm. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.

[8] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning, 2018.

[9] Electric Nation. Electric nation customer trial report. `https://eatechnology.com/media/girhcnsc/electric-nation-customer-trial-report.pdf`. Accessed 21 July 2023.

[10] Himanshu Grover, Lokesh Panwar, Ashu Verma, B.K. Panigrahi, and T.S. Bhatti. A multi-head convolutional neural network based non-intrusive load monitoring algorithm under dynamic grid voltage conditions. *Sustainable Energy, Grids and Networks*, 32:100938, 2022.

[11] G.W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[13] Michele Jamrozik, Vincent Gaudillière, Mohamed Adel Musallam, and Djamila Aouada. Space debris: Are deep learning-based image enhancements part of the solution?, 2023.

[14] Jack Kelly and William Knottenbelt. Neural NILM. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, nov 2015.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[16] Jinghui Liu, Daniel Capurro, Anthony Nguyen, and Karin Verspoor. “note bloat” impacts deep learning-based nlp models for clinical prediction tasks. *Journal of Biomedical Informatics*, 133:104149, 2022.

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2015.

[18] Sparsh Mittal. A survey on modeling and improving reliability of dnn algorithms and accelerators. *Journal of Systems Architecture*, 104:101689, 2020.

[19] Nhat Truong Pham, Duc Ngoc Minh Dang, and Sy Dzung Nguyen. Hybrid data augmentation and deep attention-based dilated convolutional-recurrent neural networks for speech emotion recognition, 2021.

[20] John Piggott and Alan Woodland. Editors' introduction. volume 1 of *Handbook of the Economics of Population Aging*, pages xvii–xx. North-Holland, 2016.

[21] Michael Pullinger, Jonathan Kilgour, Nicholas Goddard, Sarah Darby, Peter Horton, and Timothy J. Foxon. The ideal household energy dataset, electricity, gas, contextual sensor data and survey data for 255 uk homes. *Scientific Data*, 8(1):146, 2021.

[22] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[23] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks, 2014.

[24] Efe İSa Tezde and Eray Yildiz. A comprehensive survey for non-intrusive load monitoring. *Turkish Journal of Electrical Engineering and Computer Sciences*, 30(4):1162–1186, 2022.

[25] Chao Tu, Wanjun Liu, Wentao Jiang, and Linlin Zhao. Hyperspectral image classification based on residual dense and dilated convolution. *Infrared Physics & Technology*, 131:104706, 2023.

[26] Hu Yuan, Peng Zhou, and Dequn Zhou. What is low-carbon development? a conceptual analysis. *Energy Procedia*, 5:1706–1712, 2011. 2010 International Conference on Energy, Environment and Development - ICEED2010.

[27] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks, 2013.

[28] Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for nonintrusive load monitoring, 2017.

[29] Hongshan Zhao, Xihui Yan, and Hui Ren. Quantifying flexibility of residential electric vehicle charging loads using non-intrusive load extracting algorithm in demand response. *Sustainable Cities and Society*, 50:101664, 2019.