TRANSFER LEARNING METHODS FOR ENERGY DISAGGREGATION

Rohit Kumar
School of Computing
Newcastle University
Newcastle
r.kumar4@newcastle.ac.uk

Wanqing Zhao
Senior Lecturer in Data Science
Newcastle University
Newcastle
wanqing.zhao@newcastle.ac.uk

ABSTRACT

In the pursuit of global sustainability, Energy Disaggregation, known as Non-Intrusive Load Monitoring (NILM), emerges as a critical tool to optimize energy consumption. NILM's core lies in unraveling individual appliance insights from aggregated meter data, offering homes intelligent decision-making abilities. Beyond customers, the benefits extend to grid operators who may better energy distribution strategies, discover irregularities, and drive targeted energy-saving initiatives. Rooted on George Hart's pioneering work, present NILM blends signal processing and machine learning methodologies. These include probabilistic models such as factorial hidden Markov models and neural networks, such as Long Short-Term Memory Networks (LSTMs) and denoising autoencoders. These solutions excel in unraveling sophisticated energy consumption patterns, tackling challenges peculiar to NILM, especially the "unidentifiability" dilemma. Transfer learning augments NILM's capabilities by modifying models across diverse energy consumption situations. With energy consumption habits that differ across geographical locations and households, transfer learning bridges these inconsistencies, giving considerable adaptability. In conclusion, the amalgamation of deep learning approaches, inclusive of transfer learning, within the area of non-intrusive load monitoring shows promise. It provides precise and efficient energy disaggregation by applying pre-trained models and customizing them to the unique NILM circumstance. This trend not only enhances rigorous energy management but also accelerates the change to sustainable practices by cutting carbon emissions and developing energy-conscious behavior.

1 Introduction

There has never been a greater urgent need for sustainability and energy efficiency on a global scale. The desire to comprehend and optimize energy consumption, particularly in domestic settings, is at the forefront of this movement. There is an urgent need for creative solutions as the globe struggles to address the growing challenges posed by climate change, which are worsened by rising carbon emissions [1]. In this environment, Energy Disaggregation—also known as Non-Intrusive Load Monitoring (NILM)—has arisen as a guiding light, suggesting a route to more sophisticated energy management and conservation. Energy Disaggregation looks into the complex process of determining the energy usage of each individual appliance from aggregated data obtained from a single meter. The appeal of this method rests in its capacity to provide households with a detailed overview of their energy usage [2][3]. Such detailed information can be revolutionary, giving consumers the information they need to make wise decisions and encouraging them to act in an energy-efficient manner. Additionally, on a larger scale, grid operators can benefit from the information gained through NILM by optimizing energy distribution, spotting broken equipment, and even facilitating targeted energy-saving initiatives.

The foundational contributions made by George Hart in the 1980s are responsible for the development of NILM as a study area [4]. Hart's innovative work established the framework by establishing the idea of a "signature taxonomy," a collection of characteristics that could be condensed to identify certain appliance use patterns. While Hart's early research primarily focused on transitions between appliance stable states, technological improvements over the years have led to a significant expansion of NILM's application domain [5]. The development of deep learning and its subsequent growth has had a substantial impact on the direction of NILM research[6]. Deep neural architectures have been modified and adapted for energy disaggregation problems, drawing on innovations in related fields like image recognition and natural language processing. A majority of current approaches either utilize the notion of signal processing that explicitly resorts to the properties of appliances , or employ machine learning methods in supervised and unsupervised ways[7][8][9]. Specifically, many machine learning methods model the energy consumption of appliances by unsupervised probabilistic approaches such as factorial hidden Markov models (FHMM) [10] and its variants, other methods also deploy machine learning techniques such as sparse coding, matrix factorization, and k-nearest neighbor (k-NN) to separate the energy signal. Long Short-Term Memory Networks (LSTMs) and denoising autoencoders, to mention a couple, have demonstrated exceptional skill in figuring out intricate energy use patterns. They are particularly skilled at handling the complexities of NILM, notably difficulties like the "unidentifiability" conundrum, thanks to their agility and dynamism .

Transfer learning is a further aspect that deepens the NILM ecosystem. The capacity to adapt models across domains is crucial given the inherent heterogeneity in energy consumption patterns across geographical terrains and different family dynamics[11]. This model of using knowledge from one field to improve performance in another has enormous potential, particularly when traversing the complex databases that characterize global energy usage.

In conclusion, the utilization of deep learning techniques, including transfer learning, in non-intrusive load monitoring holds great potential for addressing the challenges of accurate and efficient energy disaggregation. The ability to leverage pre-trained models and adapt them to the specific task of NILM enables the development of models with improved accuracy, generalization capabilities, and reduced training time. This research direction facilitates more precise energy management and supports the transition toward sustainable and energy-efficient practices[12]

2 Background

Traditional machine learning involves building models from scratch, which frequently necessitates a significant amount of labeled data and computer power. This method delivers a paradigm change with transfer learning. It includes applying knowledge from a model that has already been trained on a similar task to a new task, which frequently results in improved performance with less data[13]. The guiding premise behind this is the conviction that knowledge gained in one field can improve performance in a related field.

2.1 Application and Benefits of Transfer Learning

The difficulties with NILM come in many forms. Appliances have distinctive energy consumption patterns, but when several devices are in use at once, these patterns might be obscured1. Additionally, these patterns vary depending on the brand, usage, wear and tear, and other environmental conditions[14]. It takes a lot of processing power to train models to recognize and precisely disaggregate these patterns. Transfer learning can assist in getting improved disaggregation accuracy with fewer data points by utilizing models that have already been trained on large datasets. Convolutional neural networks (CNNs), which are typically used for image recognition, can be modified for NILM, for example. They are suitable for detecting appliance operations because of their capacity to identify features like the abrupt surge of 1000 watts. The horizon is further broadened by the combination of transfer learning and generative adversarial networks (GANs). When paired with transfer learning, GANs' ability to provide synthetic data on energy use can produce models that are reliable and precise.

The possible increase in accuracy is the main advantage of incorporating transfer learning into NILM. Transfer learning can be a big help, especially in situations when gathering a lot of labeled data is difficult. For instance, models like RNNs and autoencoders have demonstrated respectable outcomes when paired with transfer learning principles. Additionally, Gans's ability to generate synthetic data can help with data scarcity problems by serving as a link between intensive pre-training and fine-tuning on particular, smaller datasets.

2.2 Previous Work

The technical intricacies of the Non-Intrusive Load Monitoring (NILM) model are covered in this section. We'll take a tour of the various machine learning approaches used for NILM, roughly classifying them as supervised and unsupervised learning strategies[13].

At its core, NILM uses just the agglomerated data from a central meter to disaggregate and comprehend the power usage of individual appliances within a home. Imagine a situation where the cumulative consumption of all active appliances at a specific moment, t, is represented by the total power readings, represented by Y(t)[13, 15]. This connection can be mathematically stated as:

$$Y(t) = \sum_{i=1}^{I} X_i(t) + e(t) \tag{1}$$

Here, $Xi(t)$ denotes the power consumption of the i^{th} appliance at time t, with I being the total number of appliances. The term $e(t)$ encapsulates the inherent noise in the model. Typically, this noise follows a Gaussian distribution centered around 0, with a variance $\sigma^2(t)$. Given our observations of the mains Y over a period T, the challenge lies in deducing the power consumption of each individual appliance, represented by Xi, over the same period. This scenario is emblematic of the Single-channel Blind Source Separation dilemma: the task of discerning multiple sources from a singular observation. A notable characteristic of this problem is its non-identifiability[16].

Numerous methods have been developed over time to overcome NILM's difficulties, particularly the identifiability barrier[17]. The general consensus among researchers is that incorporating domain-specific knowledge into the model can greatly lessen this problem. In the past, a variety of approaches have been put forth, including clustering algorithms, matrix factorization, and signal processing methods that make use of appliance-specific properties for disaggregation[18].

Unsupervised and supervised learning are the two most common machine learning methodologies that stand out in the huge field of NILM solutions. In the sections that follow, we will go more deeply into the specifics of these methods and examine the algorithms that have played a crucial role in determining the course of NILM research[19].

Numerous studies have been done to investigate the use of transfer learning techniques for non-intrusive load monitoring. The exact breakdown of household appliance energy use using these investigations has yielded encouraging findings. For instance, Ma et al. established a deep learning framework for energy disaggregation in another study utilizing transfer learning[20]. On a sizable dataset of labeled energy consumption data, the researchers utilized a pre-trained deep neural network model, then they refined it on a smaller sample for a particular home. The results showed that, even in the presence of noise and fluctuations in the household environment, the transfer learning approach obtained a high level of accuracy in disaggregating the energy consumption of specific appliances.

In another study, Chaoyun Zhang[6] proposed an innovative approach to energy disaggregation, rooted in sequence-to-point learning, which represents a marked departure from the established sequence-to-sequence methodologies championed by researchers like Kelly and Knottenbelt[21]. Kelly's deep learning-based approach sought a nonlinear regression between sequences of main and appliance readings. Mathematically, Kelly's approach could be represented as:

$$Y(t) = \sum_{i=1}^{I} Xi(t) + e(t) \tag{2}$$

with a loss function expressed as:

$$L_s = \sum_{t=1}^{T-W+1} \log p(X_{t:t+W-1}|Y_{t:t+W-1}, \theta_s) \tag{3}$$

In contrast, Zhang's proposition streamlines the neural network's target. By emphasizing the prediction of the midpoint element of a given window, Zhang introduces a computational elegance to the process. This sequence-to-point learning technique, which has parallels in speech and image distribution modeling, operates on the foundational assumption that the state of an appliance at a specified midpoint is intrinsically connected to the mains' readings preceding and succeeding at that point. This is represented by:

$$x_\tau = F_p(Y_{t:t+W-1}) + \epsilon \tag{4}$$

where the loss function is articulated as:

$$L_p = \sum_{t=1}^{T-W+1} \log p(x_\tau|Y_{t:t+W-1}, \theta_p) \tag{5}$$

Such an approach is intuitively grounded in real-world dynamics. The behavior of an appliance at any moment is frequently a function of its immediate past and anticipated future actions. Zhang's seq2point model offers distinct advantages, notably in terms of prediction precision. By yielding a direct prediction for each point x_t, rather than an averaged prediction across multiple windows, it potentially enhances the granularity and accuracy of the model's insights. Incorporating change points or edges in the mains as influential features for the neural network further accentuates the model's discernment capabilities, optimizing its ability to detect nuanced shifts in appliance states. In summation, Zhang's sequence-to-point learning strategy for NILM is not just an innovative departure but potentially a more efficient and precise method for energy disaggregation. The intricate interplay between mains' readings and their consequential influence on specific appliance midpoints underscores a promising avenue for further exploration in the realm of energy disaggregation research.

These studies demonstrate the possibility of non-intrusive load monitoring for household energy disaggregation to increase accuracy and efficiency using transfer learning techniques. The accuracy and effectiveness of non-intrusive load monitoring for residential energy disaggregation have been demonstrated to have tremendous potential for improvement. Transfer learning enables more precise and effective energy disaggregation by using knowledge gained from one dataset and transferring it to another.

3 Methodology

3.1 Datasets

The disaggregation of energy is possible using a variety of open-source datasets. These datasets were compiled from residences located in several nations. Sensors were deployed inside these buildings to monitor active power, and some of these sensors also collected additional information like reactive power, current, and voltage. The goal of Non-Intrusive Load Monitoring (NILM) is to make use of the active power data. However, the main difference between both datasets can be seen in the sample frequencies. As a result, preprocessing measures are required to synchronize the readings before NILM algorithms are applied to the data.

The kettle, microwave, fridge, dishwasher, and washing machine are five appliances that are frequently thought of for disaggregation in the literature, as mentioned in and. Three residential electrical datasets—REFIT, UK-DALE, and REDD will be used in our own studies.

REFIT:

From 2013 to 2015, 20 buildings in the Loughborough region of England provided data for the REFIT dataset. This dataset includes active power readings for both total mains usage and specific appliances, which are captured every 8 seconds. The largest of the three datasets being examined is REFIT. As a result, it provides the framework for deep learning model training. The justification is that a large number of electricity readings can make it easier for trained deep-learning techniques to generalize, enabling their use in new, previously unexplored households. Notably, a thorough inspection of the information before model training revealed that homes 13 and 21 used solar panel energy generation, making them irrelevant to our current research.

UK-DALE

Data from 5 buildings in the UK are included in the UK-DALE (UK Domestic Appliance-Level Electricity) dataset, which covers the years 2013 to 2015. The sampling intervals for appliances and the mains were set at 1 second and 6 seconds, respectively. Please see for more comprehensive information and statistical analysis.

3.2 NILMTK

NILMTK was created in April 2014 when this open-source toolkit first appeared, revolutionizing the analysis of both datasets and well-known techniques. This adaptable tool also functions as a platform that unifies diverse datasets and methodologies[22]. A shared data format called NILMTK-DF brings the clever representation of datasets within NILMTK to life. Based on the hierarchical data format (HDF5), this format intelligently divides data into digestible chunks that are loaded into operational memory. NILMTK-DF carefully aligns with the NILM-metadata architecture and expands its reach to encompass important metadata in addition to containing data about energy. It also accepts a variety of sensor-derived data, such as inputs for gas, water, and temperature.[23]

A collection of software elements built on the Python programming language serves as the toolkit's heart. These shining stars include (a) converters and parsers ready to take on datasets, (b) diagnostic tools skillful at revealing gaps and dropout rates within datasets, (c) the statistical powerhouse skilled at dissecting elements like the proportion of submetered energy, and (d) data preprocessing wizards skilled at down-sampling and voltage normalization. In addition to setting the stage for a thorough evaluation, NILMTK introduces two benchmark disaggregation techniques. Combinatorial optimization (CO), invented by Hart in 1985, coexists on stage with the mysterious factorial hidden

Markov models (FHMM), created initially by Kim and improved by Kolter Matthew in 2011. Performance measures like EE, ETEA, and FTEAC can be heard echoing through NILMTK's hallways[24].

The developers of this toolset set out on an evaluation expedition with detailed dataset analysis and exacting energy disaggregation benchmarks. In these benchmark tests, six datasets—REDD, Smart*, PSRI (formerly known as DataPort), AMPds, iAWE, and the UK-Dale—all studied at the granularity of 1-minute intervals—were compared against the two default algorithms. The results showed that FHMM outperformed CO in datasets including REDD, Smart*, and AMPds, whereas in the other competitions, the algorithms performed similarly.

Looking into the digital sphere, the NILMTK source code is revealed on the large Github platform. The project currently proudly bears the footprints of contributions from 17 different persons. The most recent orchestration of the code update symphony took place in the month of March of the year 2018. This lively cadence conveys that the enterprise not only keeps its attraction but also continues to capture the attention of a community united in support of its cause over time.

3.3 Experiment

3.3.1 Appliance Selection

The experimentation comprised five unique target appliances: the refrigerator, washing machine, dishwasher, kettle, and microwave. This selection was chosen, given the presence of each appliance in a minimum of three residences within the UK-DALE dataset. This decision enabled the training of networks across several dwellings and subsequent testing on diverse residences. These appliances were chosen according to their high energy consumption contribution. Moreover, they produce varied power signatures, ranging from the easy on/off pattern of a kettle to the complicated pattern presented by a washing machine (depicted in Figure 1), thus presenting a broad spectrum for investigation. Smaller-scale devices such as gaming consoles and phone chargers provide issues for numerous NILM algorithms due to their low individual impact on aggregate power demand, frequently disguised inside background noise. Despite their small individual energy usage, modern residences tend to host a plethora of such devices, together yielding a substantial energy impact. Thus, the idea of identifying these "small" appliances using NILM is an exciting path. While the current analysis hasn't studied the performance of neural networks on these "small" appliances, such exploration is planned for the future.

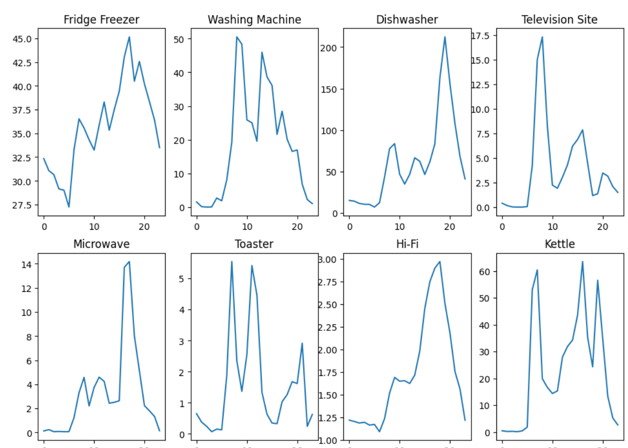


Figure 1: Load Signatures

3.3.2 Capturing Appliance Activations

Appliance activations were recovered utilizing the get_activations() method within NILMTK. The approach for each appliance is given in Table 1 . For simpler devices like toasters, activations are identified through sequential samples reaching a predefined power threshold. Subsequently, activations with durations below a designated threshold are eliminated to eliminate transitory spikes. With more sophisticated appliances like washing machines, which at times exhibit power demand below the threshold during a cycle, NILMTK adjusts for these brief sub-threshold periods. Figure **??** depicts the cluster of activation functions of the appliances.

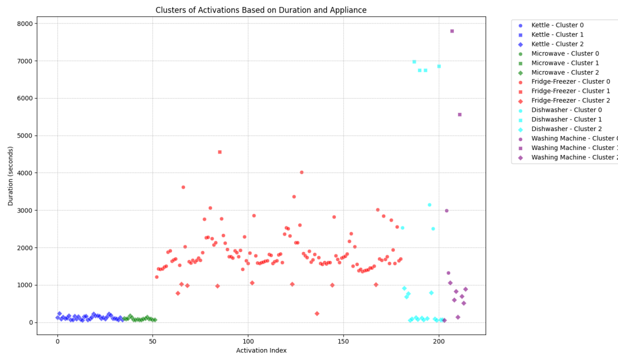


Figure 2: Clusters of Activations of Appliances based on Duration

Table 1: Appliance Config

Appliance	Max On power (watts)	Min. on power (watts)	Min. off duration (secs)	Appliance power duration (secs)
Kettle	3100	2000	12	0
Fridge	300	50	60	12
Washing m.	2500	20	1800	160
Microwave	3000	200	12	30
Dishwasher	2500	10	1800	1800

3.3.3 Selection of Real Aggregate Data Windows

The method begins by detecting all activations of the target appliance within the home's submeter data for that specific appliance. Subsequently, for each training instance, a random decision is taken with a 50% probability about the inclusion of the target appliance. If exclusion is chosen, a random aggregate data window devoid of activations of the target appliance is picked. Conversely, if inclusion is agreed upon, a target appliance activation is randomly picked and positioned within the data window meant for the neural network's target. This technique assures that the full activation is contained within the data window given to the network unless the window length imposes limits.

3.3.4 Generation of Synthetic Aggregate Data

Synthetic aggregate data production commences with the aggregation of appliance activations across all training homes for the five target appliances: kettle, washing machine, dishwasher, microwave, and refrigerator. Creating a single sequence of synthetic data includes initializing two vectors with zeros, one for the network's input and the other for the target. The vector lengths define the width of the data window encountered by the network. The synthetic data sequence evolves by sequential judgments for each appliance class, deciding the inclusion of an activation of that class in the training sequence. Probabilities are set at 50% for the target appliance and 25% for each "distractor" appliance. Chosen appliance activations are randomly positioned inside the input vector. Distractor appliances can occur anywhere in the series, whereas target appliance activations are ensured to be fully contained within the sequence.

It's vital to realize that this rather simplified approach of synthesizing aggregate data disregards different complexity found in genuine aggregate data. Real-world patterns, such as the co-occurrence of a kettle and a toaster activating within a short time frame, are not captured by our crude model. Employing a more advanced simulator could potentially boost the effectiveness of deep neural networks in energy disaggregation.

3.3.5 Standardization

The best learning of neural networks often needs input data with a mean of zero. To do this, the mean of each sequence is deducted from the series, resulting in a mean of zero for each. Furthermore, every input sequence is divided by the standard deviation of a random sample from the training set. To preserve scaling, sequences are not separated by their individual standard deviations, as this scaling has significance for NILM.

While imposing zero mean on each sequence leads to information loss, particularly pertinent to NILM algorithms like combinatorial optimization and factorial hidden Markov models, preliminary experiments indicate that independent centering of each sequence might enhance generalization within neural networks. Striking the proper balance could potentially provide the network information with absolute power while permitting effective generalization. A noteworthy feature of training sequences with independent centering is their immunity from vampire loads—persistent loads that are continually engaged. To maintain consistency, target values are divided by a predetermined "maximum power demand" particular to each appliance, effectively normalizing target power demand to the range of [0, 1].

4 Network Architecture

4.1 1D ResNet-50 Architecture

The Residual Network (ResNet) architecture was first developed to overcome the vanishing gradient issue in deep neural networks by inserting skip connections, enabling the network to learn identity functions. In this work, we have adopted the ResNet-50 architecture, which has 50 layers and tuned it for 1-dimensional (1D) data.

The main building block of our tuned ResNet-50 architecture comprises a sequence of 1D convolutional layers followed by batch normalization and activation algorithms. Specifically:

Identity Block: This block has no convolution layer at the shortcut. Each identity block consists of:

- A 1D convolution with a filter size of 1, which is followed by batch normalization and ReLU activation.
- Another 1D convolution with a defined kernel size (usually 3) and padding, followed by batch normalization and ReLU activation.
- A third 1D convolution with a filter size of 1. This is then added to the original input tensor to generate the residual connection.
- The product then undergoes another ReLU activation.

Convolution Block: This block has a convolution layer at the shortcut. Each convolution block consists of:

- A 1D convolution with a filter size of 1, followed by batch normalization and ReLU activation.
- Another 1D convolution with a chosen kernel size (usually 3) with padding, followed by batch normalization and ReLU activation.
- A third 1D convolution with a filter size of 1.
- Parallelly, the input tensor also experiences a 1D convolution with a filter size of 1, which is subsequently added to the main route to generate the residual connection.
- The product then undergoes another ReLU activation.

The above blocks are layered progressively with increasing filter sizes to produce the depth of the ResNet-50 architecture.

At the end of the network:

- An average pooling procedure with a window size of 7 is applied.
- If the model is configured to include the top, it is flattened and may optionally pass through thick layers with specified activations. Dropout may also be used between these thick layers for regularization. The last thick layer delivers the forecasts.

An average pooling procedure with a window size of 7 is applied. If the model is configured to include the top, it is flattened and may optionally pass through thick layers with specified activations. Dropout may also be used between these thick layers for regularization. The last thick layer delivers the forecasts.

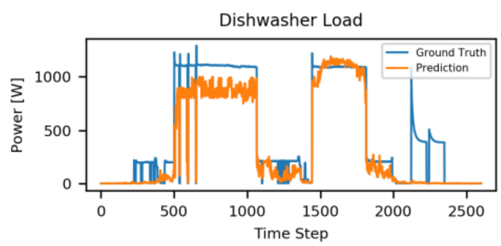
The whole architecture is compiled using a mean squared error (MSE) loss and the Adam optimizer. The measure used to assess the model's performance is accuracy

5 Result and Evaluation

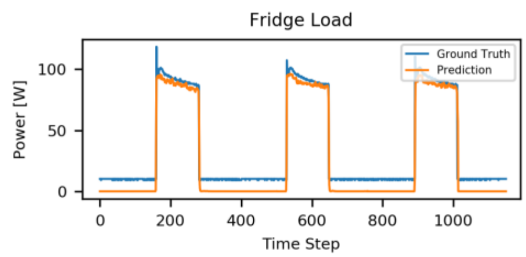
The ResNet-50 CNN model was originally used for a rather little dataset, which is the UK-DALE dataset. During its initial assessment, the model was solely evaluated inside the same domains as its training data. This required training on the UK-DALE dataset and then analyzing further UK-DALE data. This was done as part of initial testing and trial experiments to understand the model and the dataset. Then a similar method was followed with the REFIT dataset.

However, the issue of whether the proposed ResNet-50 CNN model's applicability might expand to new domains remained unclear. A major difficulty with deep neural networks is their tendency to overfit to a specific domain, especially when the training data distribution greatly varies from that contained in the testing data. In the context of Non-Intrusive Load Monitoring (NILM), home data is received from numerous nations, meaning that appliances could display dramatically divergent patterns across different locations. Due to this reason instead of focusing on several prominent datasets such as REDD, and DEDDIAG, which are from different nations, we stick to the two popular UK datasets, UK-DALE and REFIT. Even within a single nation, changes in appliance behavior across different households might further increase the variance. Our next experiments will comprehensively evaluate the transferability of the ResNet-50 CNN model and highlight its potential for successful deployment in transfer learning settings.

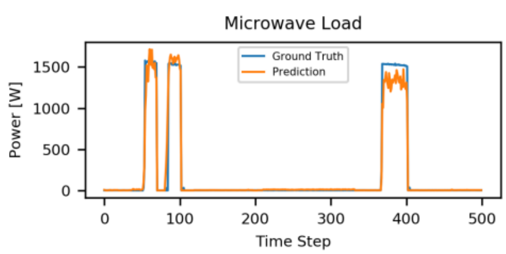
REFIT has a considerably greater dataset compared to UK-DALE, resulting in a more effective generalization of the model. On the contrary, UK-DALE lacks a comparable domain to REFIT, presumably explaining why the usage of transfer learning did not deliver performance gains for the ResNet-50 CNN model. Table 2 demonstrates that the ResNet-50 CNN model described in the research outperformed the other model AFHMM. The ResNet-50 CNN model was able to lower the MAE and SAE values for all the appliances compared to the reference AFHMM model. The ResNet-50 CNN lowered the average MAE of the five appliances by 83.63% and SAE by 77.76%. It also reduces MAE by 24.72% with regard to seq2seq(MingjunZhong). Figure 3 depicts the Energy Disaggregation prediction of the appliances on the REFIT dataset.



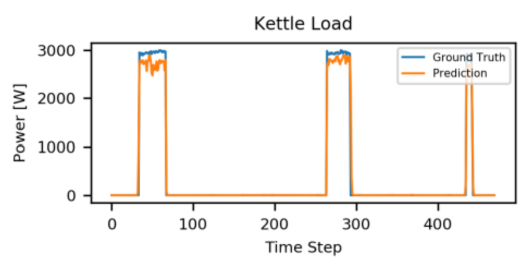
(a) prediction using ResNet-50 model trained on Dishwasher



(b) prediction using ResNet-50 model trained on Fridge



(c) prediction using ResNet-50 model trained on Microwave



(d) prediction using ResNet-50 model trained on Kettle

Figure 3: Disaggregation Results for REFIT

Appliance Transfer Learning:

The CNN model is trained on the dishwasher and will be utilized for all the other appliances. The CNN layers were trained using REFIT. Table 2 displays the results of ATL, where the CNN layers were trained on REFIT using the dishwasher, and evaluated on REFIT. The outcomes are equivalent to the usual training. There is not a significant difference between the predictions provided by normal training compared to the prediction via appliance transfer learning which is highlighted by the small difference in the MAE and SAE values. We also deduce that merely by training the ResNet-50 CNN model on a big dishwashing dataset we can apply it to other appliances and achieve acceptable predictions because they are also employing similar sorts of load signatures for NILM.

Cross-Domain Transfer Learning:

The notion behind cross-domain transfer learning is that information collected from one domain (source) may be applied to enhance the performance of another, however related, domain (target). In our situation, the source domain is REFIT, and the destination domain is UK-DALE. For our model, CTL did not produce very remarkable outcomes which can be attributed to issues like:

- Model Architecture and Hyperparameters: The architecture and hyperparameters selected could be optimum for REFIT but not for UK-DALE.
- Training technique: The fine-tuning technique and the training length could have influenced the transfer learning outcomes.
- Feature Extractors: The early layers of a neural network operate as feature extractors. If the characteristics collected from the REFIT dataset aren't representative or relevant to the UK-DALE dataset, performance might suffer.

Table 2: Appliance Transfer Learning for Sequence-to-Point Learning

	AFHMM		CNN (Trained on REFIT tested on REFIT)		CNN (Trained on REFIT tested on REFIT(ATL))		CNN (Trained on REFIT tested on UK-DALE(CTL))	
Appliance	MAE	SAE	MAE	SAE	MAE	SAE	MAE	SAE
Kettle	47.38	1.06	6.110	0.132	12.650	0.050	24.492	0.105
Microwave	21.18	1.04	11.120	0.166	13.395	0.065	26.380	0.250
Fridge	42.35	0.98	22.440	0.400	21.560	0.404	30.560	0.660
Dishwasher	199.84	4.50	11.260	0.240	12.260	0.610	21.324	0.998
Washingmachine.	103.24	8.28	16.990	2.590	16.950	2.585	27.852	3.990
Avg	82.79	3.17	13.548	0.705	15.363	0.277	26.126	1.200

6 Conclusion

Conclusively, this research goes into the area of transfer learning in the context of NILM, giving light to the possibility of using ResNet-50 CNN models as an alternative to the usual seq2point strategy. Unlike earlier initiatives constrained inside solitary domains, our technique navigates the nuances of training within one context and testing within another. This strategy focuses on the underlying concept that the key characteristics retrieved by ResNet-50 CNN networks display consistency across varied equipment and data domains.

The benefits gathered from using this transfer learning technique carry particular importance. Principally, it traces a potential trajectory toward obtaining oracle-level NILM capabilities, exemplified by a solitary model capable of understanding a large range of home appliances globally. Additionally, this paradigm carries the potential to control the sensor proliferation for specific appliances, leading to judicious cost allocations. Moreover, the strategic implementation of transfer learning leads to huge computing savings as pre-trained ResNet-50 CNN models smoothly adapt and excel across a spectrum of appliances and domains.

In short, this study travels beyond the constraints of traditional approaches, arguing for the deployment of ResNet-50 CNN models in the transfer learning paradigm, ushering in new dimensions of accuracy, flexibility, and computing resource optimization within the area of NILM.

7 Future Work

In light of the presented findings, several intriguing avenues for future exploration emerge, which have the potential to enhance the efficacy and generalizability of transfer learning in the context of NILM:

- Incorporation of Auxiliary Data Sources: To deepen the transfer learning process, the incorporation of additional context-rich data sources, such as weather information, occupancy data, and appliance metadata, offers potential. These extra datasets may give useful insights into the underlying patterns of energy usage, thereby increasing the accuracy of disaggregation. Developing efficient ways to smoothly integrate these varied data sources into the transfer learning framework might greatly increase the model's flexibility.
- Enhanced Integration of UK-DALE Dataset: While the UK-DALE dataset supplies a considerable amount of data from various families, its restricted representation across a few dwellings offers issues in generalization. Addressing this restriction, future study might require increasing the dataset by including additional households, preferably spanning hundreds or even thousands. This would create a greater spectrum of appliance variants, helping neural networks to grasp the vast range of consumption patterns observed in various environments.
- Exploration of Larger Appliance Variability: Building upon the insights garnered from the present research, future initiatives might dive further into understanding and adapting the nuances of appliance behavior.

Research efforts might concentrate on types of appliances that display considerable variety in their consumption patterns. By training on increasing numbers of cases within these classes, neural networks might achieve a heightened capacity to generalize across the various variations typical of such appliances.

References

[1] Cynthia Rosenzweig, David Karoly, Marta Vicarelli, Peter Neofotis, Qigang Wu, Gino Casassa, Annette Menzel, Terry L Root, Nicole Estrella, Bernard Seguin, Piotr Tryjanowski, Chunzhen Liu, Samuel Rawlins, and Anton Imeson. Attributing physical and biological impacts to anthropogenic climate change. *Nature*, 453(7193):353—357, May 2008.

[2] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, 2012.

[3] Roberto Bonfigli, Stefano Squartini, Marco Fagiani, and Francesco Piazza. Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview. *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, pages 1175–1180, 2015.

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

[5] *21st IEEE International Conference on Electronics, Circuits and Systems, ICECS 2014, Marseille, France, December 7-10, 2014*. IEEE, 2014.

[6] Chaoyun Zhang, Mingjun Zhong, Zong-Hui Wang, Nigel H. Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for nonintrusive load monitoring. In *AAAI Conference on Artificial Intelligence*, 2016.

[7] A L Wang, Binyang Chen, C. Gang Wang, and Dong Hua. Non-intrusive load monitoring algorithm based on features of v–i trajectory. *Electric Power Systems Research*, 157:134–144, 2018.

[8] Shengli Du, Mingchao Li, Shuai Han, Jonathan Shi, and Heng Li. Multi-pattern data mining and recognition of primary electric appliances from single non-intrusive load monitoring data. *Energies*, 12:992, 03 2019.

[9] Mario E. Berges, Ethan Goldman, H. Scott Matthews, Lucio Soibelman, and Kyle D. Anderson. User-centered nonintrusive electricity load monitoring for residential buildings. *J. Comput. Civ. Eng.*, 25:471–480, 2011.

[10] Mingjun Zhong, Nigel Goddard, and Charles Sutton. Signal aggregate constraints in additive factorial hmms, with application to energy disaggregation. volume 4, 12 2014.

[11] Zhanhong Jiang and Young M. Lee. Deep transfer learning for thermal dynamics modeling in smart buildings. *2019 IEEE International Conference on Big Data (Big Data)*, pages 2033–2037, 2019.

[12] Awadelrahman M. A. Ahmed, Yan Zhang, and Frank Eliassen. Generative adversarial networks and transfer learning for non-intrusive load monitoring in smart grids. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–7, 2020.

[13] Michele D'Incecco, Stefano Squartini, and Mingjun Zhong. Transfer learning for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*, 11(2):1419–1429, 2020.

[14] Fayiz Alfaverh, Mouloud Denai, and Yichuang Sun. Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management. *IEEE Access*, PP:1–1, 02 2020.

[15] Christoph Klemenjak, Stephen Makonin, and Wilfried Elmenreich. Towards comparability in non-intrusive load monitoring: On data and performance evaluation. In *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2020.

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[17] Antoine Langevin, Marc-André Carbonneau, Mohamed Cheriet, and Ghyslain Gagnon. Energy disaggregation using variational autoencoders, 03 2021.

[18] Kaibin Bao, Kanan Ibrahimov, Martin Wagner, and Hartmut Schmeck. Enhancing neural non-intrusive load monitoring with generative adversarial networks. *Energy Informatics*, 1, 10 2018.

[19] Thi-Thu-Huong Le and Howon Kim. Non-intrusive load monitoring based on novel transient signal in household appliances with low sampling rate. *Energies*, 11(12), 2018.

[20] R. Ramadan, Qi Huang, Olusola Bamisile, Amr S. Zalhaf, Karar Mahmoud, Matti Lehtonen, and Mohamed M. F. Darwish. Towards energy-efficient smart homes via precise nonintrusive load disaggregation based on hybrid ann–pso. *Energy Science & Engineering*, 11(7):2535–2551, 2023.

[21] Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. 11 2015.

[22] Seyed Mostafa Tabatabaei, Scott Dick, and Wilsun Xu. Toward non-intrusive load monitoring via multi-label classification. *IEEE Transactions on Smart Grid*, 8(1):26–40, 2017.

[23] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. Nilmtk: An open source toolkit for non-intrusive load monitoring. 07 2014.

[24] Amir Kavousian, Ram Rajagopal, and Martin Fischer. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, 55:184–194, 06 2013.

8 Appendices

8.1 Evaluation Metrics Used

Evaluation of performance is carried out using two metrics. Let x_t represent the actual values, and x_t denote predictions for an appliance at time t. To quantify power-related errors at each time step, the mean absolute error (MAE) is utilized:

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

The MAE mitigates the impact of outliers, such as isolated predictions, ensuring accuracy across the entire dataset. For assessing the aggregate energy error over a specific period, like a day, the normalized signal aggregate error (SAE) is employed:

$$SAE = \frac{|r - r|}{r}$$

Here, r and r denote the actual and predicted total energy consumption of an appliance, respectively. This measure is valuable as it indicates the accuracy of daily power consumption even when individual timestep predictions might be less precise.