IMG-NILM: A Deep learning NILM approach using energy heatmaps

Jonah Edmonds
University of Bristol, UK
je17667@bristol.ac.uk

Zahraa S. Abdallah
University of Bristol, UK
zahraa.abdallah@bristol.ac.uk

ABSTRACT
Energy disaggregation estimates appliance-by-appliance electricity consumption from a single meter that measures the whole home’s electricity demand. Compared with intrusive load monitoring, NILM (Non-intrusive load monitoring) is low-cost, easy to deploy, and flexible. In this paper, we propose a new method, coined IMG-NILM, that utilises convolutional neural networks (CNN) to disaggregate electricity data represented as images. IMG-NILM transforms time series into heatmaps with higher electricity readings portrayed as ‘hotter’ colours. The image representation is then used in CNN to detect the signature of an appliance from aggregated data. IMG-NILM is robust and flexible with consistent performance on various types of appliances; including single and multiple states. It attains a test accuracy of up to 93% on the UK-Dale dataset within a single house, where a substantial number of appliances are present. In more challenging settings where electricity data is collected from different houses, IMG-NILM attains also a very good average accuracy of 85%.

CCS CONCEPTS
• Computing methodologies → Neural networks; Temporal reasoning;
• Applied computing → Engineering;

KEYWORDS
Energy disaggregation, Neural network, NILM, Time Series, image representation

ACM Reference Format:
Jonah Edmonds and Zahraa S. Abdallah. 2023. IMG-NILM: A Deep learning NILM approach using energy heatmaps. In The 38th ACM/SIGAPP Symposium on Applied Computing (SAC ’23), March 27–March 31, 2023, Tallinn, Estonia. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3555776.3555778

1 INTRODUCTION
Intelligent use of electricity will be the most important factor in solving the world’s energy problems [2]. The ability to disaggregate the net energy into individual appliances will enhance consumption understanding which will benefit energy market stakeholders from consumers to suppliers. Applications for disaggregation include outlier detection to spot faulty or expensive appliances, simplifying costs and bills, and spotlighting temporal trends. An intrusive approach to energy disaggregation is accurate, however, it has many concerns in terms of privacy and cost. Alternatively, a non-intrusive Load Monitoring (NILM) takes the aggregated power consumption of the building and uses this information to estimate the individual appliance data. This approach is more practical, cheaper and ethically less controversial, therefore, being a much more viable and attractive prospect to consumers concerned about their privacy.

Deep neural networks (DNNs) have become popular in many domains such as image recognition, speech translation and automated cancer cell detection [10]. DNNs are also applied to NILM, yet still encounter many challenges in terms of accuracy and robustness, especially when applied across different types of appliances. This paper, therefore, utilises DNNs for NILM to address some of these challenges. The contribution of this paper is summarised as follows: 1) Transforming time series data into the form of a heatmap which preserves the time dependency whilst revealing patterns related to user behaviour, routine, weather and other time-dependent factors. 2) IMG-NILM showed very good performance across three different types of appliances including multi-state types which is harder to classify. 3) IMG-NILM proved robust when the number of appliances is increased.

2 LITERATURE REVIEW
The majority of modern approaches to NILM disaggregation utilise deep neural networks. For instance, Kelly and Knottenbelt [5] depicted a sequence-to-sequence (seq2seq) method that adapted three different DNNs to fit a NILM model in the form of a ‘long short-term memory’ (LSTM) architecture, a ‘denoising autoencoder’ (DAE) and a CNN that regresses the start time, end time and average power demand for each appliance activation. LSTMs are a form of recurrent neural network (RNN) meaning that there are feedback connections within their architecture which works as an ‘artificial memory’. The main function of an autoencoder is forced to reconstruct and reduce noise in the original data, a problem that aggregated data can cause, as is concurred by Siddiqui and Sibal [12]. LSTM is then applied to classify the data based on trained data of power signatures. LSTM showed a very good performance with appliances of two states but struggled to classify appliances which had multiple running states. Also, the method is computationally inefficient.

Gomes and Pereira [3] used RNNs with a combination of CNNs guided by a pinball loss function. A pinball loss function is implemented to take into account appliances with unpredictable power patterns and this translated into very high accuracy levels when compared to other architectures as well as other loss functions. The major drawback of this work arises when attempting to disaggregate energy data coming from two appliances with similar energy outputs. Representing time series in NILM as an image is
a reasonably new concept and therefore a few papers in the literature have specifically discussed it. Kyrkou et al. [9] transformed energy consumption time series signals into polar coordinates and then utilised the Gramian Angular Summation/Difference Field (GASF/GADF) transformations to create 2-D images. The focus of this study is transfer learning when a pre-trained VGG16 model is used for classifying appliances. This study showed that models using image representation have the potential to generalise better to new data, even when the data is from an entirely different dataset. Senaratna et al. [11] compared various image representations for NILM, namely Gramian Angular Fields (GAF), Gramian Angular Difference Field (GADF) and Recurrence Plots (RP). The results showed that Gramian Angular Difference Field (GADF) outperforms other image representations. Although these representations demonstrated good potential for image representation in NILM, they are based on correlation rather than generated directly from the energy consumption, whereas the temporal aspect is preserved. Also, most of the aforementioned methods focus on the disaggregation of one type of appliance and do not perform well across different types. Another research that uses the heatmap representation but for genetic data is discussed in [1].

3 METHODOLOGY

The method starts with an input of time series to generate a heatmap such as in Figure 1. Each pixel in the heatmap corresponds to a unit of aggregated data. For a window of size \( w \), the data is aggregated for each \( i \in w \) according to an aggregation step \( s \). For instance, in Figure 1, data is presented for a duration of a day, hence the window size; \( w \); is 24 (in hours unit). The aggregation step for this heatmap is specified as 5 seconds; \( s = 5 \). Hence, for each \( i \) in \( w \) corresponds to an hour of the day, each pixel represents an aggregated data of \( s \); 5 seconds, with a total of 720 pixels per hour (column). The data is normalised using z-score normalisation to standardise across different windows and houses. The heatmap depicts the consumption of power on a specific day with dark blue colour showing low consumption that transitions through a lighter blue and into a red colour as the power values increase. The image shows a clear increase in power consumption during evenings which could be explained by turned-on heaters, with some scattered patterns relevant to routine-related activities such as cooking or watching TV mostly in the afternoon.

![Image](image_url)

Figure 1: Heatmap depicting the total power readings for house 1 in the UK-Dale dataset

The goal of disaggregation using IMG-NILM is to differentiate between time periods when appliances were turned on/off. Hence, the classifier can decide whether a certain appliance is switched off from the heatmap of a specific window.

In convolutional neural networks (CNNs), convolutional layers tend to record the exact location of a feature, and this can reduce the generality of a model if the feature map is sensitive to the location of a pattern in the image. The input layer in IMG-NILM is in a form of a collection of heatmaps/images of size 300 X 300 for three channels; representing RGB images. The architecture is composed of three consecutive convolutional layers, each followed by a pooling layer to perform down-sampling of the spatial dimension of the input. The convolutional layers are followed by batch normalisation [4] to improve learning and Rectified Linear Unit (ReLU). The network is finalised with three fully connected layers, all of which utilise a softmax activation function. Naturally, this model is paired with a categorical cross-entropy to detect the presence or absence of an appliance from an aggregated heatmap.

4 EXPERIMENTAL ANALYSIS

For the analysis we use the UK-Dale dataset [6]. More than 13 thousand images are generated for this study across appliances. Window size \( w \) is set in all experiments to one day (24 hours) while the aggregation step \( s \) is set to be 6 seconds. A commonly used metric is the training/validation/test accuracy of the model. Accuracy is calculated by: accuracy = (TP+TN)/(TP+TN+FP+FN), Where TP is true positive, FP is false positive, TN is true negative and FN is false negative. We run our experiments with an 80:20 ratio for the training:test split. With a 20% split for the validation set of the training set. All of the models utilise the ‘adam’ optimiser with a loss rate of \( 1 \times 10^{-4} \). Household appliances are categorised into four types. We focus in this paper on the three most common types of I, II and IV.

4.1 Experiments on a single-house data

We start with data from a single house, where the model is trained to learn certain patterns that the CNN identifies and therefore return whether a certain appliance is on or off. This particular house has a large number of appliances (512) across the four types, hence the classification is challenging. In the instance of the dishwasher classes, both the training accuracy and the validation accuracy with scores of 92% and 88% respectively. The test results further back up the validity of this test with a test accuracy value of 85%. Dishwashers, a Type II appliance, have irregular but strong power signatures, using a lot of electricity for a specified length of time. Due to this, when converted into a heatmap image, the dishwasher provides a specific ‘heat signature’ which displays itself as a pattern on the image.

However, refrigerators (Type IV) are, in general, always turned on. Their power output remains somewhat constant and therefore, when this appliance is turned off, the impact on the heatmap is mainly on reducing the colour intensity of every pixel across the whole day. A CNN looking for patterns within the images will therefore struggle to find any similar ‘heat signatures’ that would identify the presence/absence of a refrigerator. Hence, the number of epochs is adjusted to 250, instead of 120 for the dishwasher, to capture the difference between appliances in the case of refrigerators and TVs. Parameter tuning is performed on the validation set.
A summary of the results across the three appliances in House 1 data is shown in Table 1. Training accuracy is very high across all appliances, however in testing, classifying refrigerators achieves the best accuracy, while TV and Dishwasher are less accurate at 86% and 85% respectively.

Table 1: Summary of IMG-NILM accuracy on UK-Dale house 1 dataset across three Types of appliances

| Appliances         | Training | Validation | Testing |
|--------------------|----------|------------|---------|
| Refrigerator (Type IV) | 98%      | 91%        | 93%     |
| TV (Type I)        | 94%      | 82%        | 86%     |
| Dishwasher (Type II) | 92%      | 88%        | 85%     |

These results show the robustness of the proposed model across different appliances demonstrating the model’s ability to learn different patterns of diverse types. Despite the large number of appliances in the aggregated images in this house (52 appliances), IMG-NILM can detect devices of different types with high accuracy. A multi-state device (dishwasher) is showing a robust performance, very close to Type I, which is a non-trivial task in NILM.

4.2 Experiments on across-houses data

In the following experiments, we look to expand on our data to include heatmaps from different houses. Adding the data from the additional houses results in a harder task with complicated heatmaps. To avoid over-fitting, a drop-out layer is added with a drop-out rate specified as 0.25 in the following experiments.

Table 2 summarises training, validation, and testing accuracy for each appliance representing the three types. As expected when working with data from across houses, validation accuracy has dropped from what has been reported in Table 1 when using data from a single house. However, IMG-NILM still attains a consistently high accuracy across different types of appliances. It is paramount therefore that the model does not overfit a particular house, especially since no two homes are the same in energy usage. One of the variables that have a major impact on the power output of a home is the brands and models of everyday appliances, with different models of the same appliance often showing different heat signatures altogether.

Table 2: Summary of the IMG-NILM accuracy of the UK-Dale dataset on different houses across three Types of appliances

| Appliance         | Training | Validation | Testing |
|-------------------|----------|------------|---------|
| Refrigerator (Type IV) | 89%      | 74%        | 84%     |
| TV (Type I)       | 93%      | 69%        | 83%     |
| Dishwasher (Type II) | 94%      | 80%        | 87%     |

Comparing our method to the literature has encountered multiple challenges. Gomes and Pereira [8] reported estimated test accuracy of 89% when attempting to predict the presence of a dishwasher. However, the design of the models in the discussed paper is specifically structured to work effectively on Type II appliances. Although their results are slightly higher than IMG-NILM (88% for dishwashers), IMG-NILM is not exclusive to one type and can be applied to various types of appliances, while still maintaining good accuracy. A relevant image-based NILM method is presented in [11]. However, this has been only tested for one house on the REDD dataset by [7], which is much smaller and has a fewer number of appliances. Also, the study looked at refrigerators only. The average validation accuracy remains approximately constant at 94.2%, 89.9%, and 82.7% for GADF, GASF, and RP, respectively. The comparable accuracy we attain with IMG-NILM on a refrigerator tested on a single house is 91% validation accuracy and 93% for testing. It is also noted that the data from this house is longer and more complex due to the large number of appliances that exist.

5 CONCLUSION AND FUTURE WORK

In this paper, we introduce IMG-NILM which adapts the CNN architecture for energy disaggregation based on images, i.e., heatmaps, generated from power consumption. The proposed approach shows an accurate and robust performance in challenging scenarios, both in a single house when many devices are operating and also across houses. IMG-NILM can be successfully used for different types of appliances with single or multiple states, constant or variant power consumption. Future work will focus on extending the model to detect multiple appliances at the same time, which is one of the main challenges in NILM. Also, we will focus on investigating the parameters for generating images and the best window for each appliance. An extensive comparison with time series-based methods and image representation methods such as GASF/GADF are a priority in future work.

REFERENCES

[1] Pierangela Bruno and Francesco Calimeri. 2019. Using heatmaps for deep learning based disease classification. In 2019 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB). IEEE, 1–7.
[2] Arpad Gellert, Adrian Florea, Ugo Fiore, Francesco Palmieri, and Paolo Zanetti. 2019. A study on forecasting electricity production and consumption in smart cities and factories. International Journal of Information Management 49 (2019), 546–556.
[3] Eduardo Gomes and Lucas Pereira. 2020. PB-NILM: Pinball guided deep non-intrusive load monitoring. IEEE Access 8 (2020), 48386–48398.
[4] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning, PMLR, 448–456.
[5] Jack Kelly and William Knottenbelt. 2015. Neural nilm: Deep neural networks applied to energy disaggregation. In Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments. 55–64.
[6] Jack Kelly and William Knottenbelt. 2015. The UK-Dale dataset. domestic appliance-level electricity demand and whole-house demand from five UK homes. 2, 150007 (2015). https://doi.org/10.1038/sdata.2015.7
[7] J Zico Kolter and Matthew J Johnson. 2011. REDD: A public data set for energy disaggregation research. In Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA, Vol. 25. 59–62.
[8] Weicong Kong, Zhao Yang Dong, Bo Wang, Junhua Zhao, and Jie Huang. 2020. A Practical Solution for Non-Intrusive Type II Load Monitoring Based on Deep Learning and Post-Processing. IEEE Transactions on Smart Grid 11, 1 (2020), 146–160.
[9] Lampri Kyrkou, Christoforos Nalmpantis, and Dimitris Vrakas. 2019. Imaging time-series for NILM. In International Conference on Engineering Applications of Neural Networks. Springer, 188–196.
[10] Wojciech Samek, Grégoire Montavon, Sebastian Lapuschkin, Christopher J Anderson, and Klaus-Robert Müller. 2021. Explaining deep neural networks and beyond: A review of methods and applications. Proc. IEEE 109, 3 (2021), 247–278.
[11] Manuri Senarathna, Madhawa Herath, Thakshila D Thilakanayake, Migara H Liyanage, and Chitral J Angammana. 2021. An Image Based Approach of Energy Signal Disaggregation Using Artificial Intelligence. In 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS). IEEE, 98–103.
[12] Abdul Siddiqui and Ananya Sibal. 2020. Energy disaggregation in smart home appliances: a deep learning approach. Energy (2020).

1153