

On the Use of Concentrated Time-Frequency Representations as Input to a Deep Convolutional Neural Network: Application to Non Intrusive Load Monitoring

Sarra Houidi, Dominique Fourer, François Auger

▶ **To cite this version:**

Sarra Houidi, Dominique Fourer, François Auger. On the Use of Concentrated Time-Frequency Representations as Input to a Deep Convolutional Neural Network: Application to Non Intrusive Load Monitoring. Entropy, 2020, 22 (9), pp.911. 10.3390/e22090911 . hal-02917749

HAL Id: hal-02917749
https://hal.science/hal-02917749

Submitted on 19 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

entropy

MDPI

Article

On the Use of Concentrated Time–Frequency Representations as Input to a Deep Convolutional Neural Network: Application to Non Intrusive Load Monitoring

Sarra Houidi [1,2], Dominique Fourer [1,*] and François Auger [2]

[1] Laboratoire IBISC (Informatique, BioInformatique, Systèmes Complexes), EA 4526, University Evry/Paris-Saclay, 91020 Evry Cédex, France; Sarra.Houidi@gmail.com
[2] Institut de Recherche en Energie Electrique de Nantes Atlantique (IREENA), EA 4642, University of Nantes, 44602 Saint-Nazaire, France; Francois.Auger@univ-nantes.fr
* Correspondence: Dominique.Fourer@univ-evry.fr

Received: 29 July 2020; Accepted: 17 August 2020 ; Published: 19 August 2020



Abstract: Since decades past, time–frequency (TF) analysis has demonstrated its capability to efficiently handle non-stationary multi-component signals which are ubiquitous in a large number of applications. TF analysis us allows to estimate physics-related meaningful parameters (e.g., F_0, group delay, etc.) and can provide sparse signal representations when a suitable tuning of the method parameters is used. On another hand, deep learning with Convolutional Neural Networks (CNN) is the current state-of-the-art approach for pattern recognition and allows us to automatically extract relevant signal features despite the fact that the trained models can suffer from a lack of interpretability. Hence, this paper proposes to combine together these two approaches to take benefit of their respective advantages and addresses non-intrusive load monitoring (NILM) which consists of identifying a home electrical appliance (HEA) from its measured energy consumption signal as a "toy" problem. This study investigates the role of the TF representation when synchrosqueezed or not, used as the input of a 2D CNN applied to a pattern recognition task. We also propose a solution for interpreting the information conveyed by the trained CNN through different neural architecture by establishing a link with our previously proposed "handcrafted" interpretable features thanks to the layer-wise relevant propagation (LRP) method. Our experiments on the publicly available PLAID dataset show excellent appliance recognition results (accuracy above 97%) using the suitable TF representation and allow an interpretation of the trained model.

Keywords: non-intrusive load monitoring (NILM); time–frequency representation (TFR); deep learning; convolutional neural network (CNN); synchrosqueezing; Layer wise relevance propagation (LRP)

1. Introduction

Real-world signals such as electrical signals can be modeled as a mixture of non-stationary components which can be disentangled using efficient methods allowing to jointly represent them in time and in frequency. Among popular time–frequency (resp. time–scale) analysis techniques, Short-time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT), which are rooted in a mature mathematical background, have shown their efficiency in a large number of applications [1,2] (eg., audio processing). However, such techniques suffer from theoretical limitations, mostly explained by the Heisenberg-Gabor uncertainty principle which can be compensated using promising post-processing methods such as reassignment and synchrosqueezing [3,4]. These methods were

proposed to sharpen a Time–Frequency (TF) representation by moving the transforms values to more accurate coordinates closer to the exact TF support of the signal, resulting in an improved readability. More particularly, the success of the synchrosqueezing transform comes from its capability to compute sharpened reversible Time–Frequency Representations (TFRs), which enable advanced processing methods for disentangling and extracting the elementary signal components (modes) [5,6].

Nowadays, to the best of our knowledge, there exists no study which investigates the relevance of such post-processing techniques when they are combined with a machine learning framework based on Convolutional Neural Network (CNN) in a realistic application scenario [7]. Moreover, representation learning is an active research subject [8] because it addresses the question of the learning model explainability and interpretability. Hence, developing understandable models is full of interest since it enables the validation of machine learning algorithms and paves the way to robust and efficient methods. This study aims at filling the gap by investigating a hybrid approach which consists of combining time–frequency signal processing techniques with Deep Neural Network (DNN) architectures. We address Non Intrusive Load Monitoring (NILM) [9] as a "toy" problem, which consists of a Home Electrical Appliance (HEA) identification from the observation of voltage and current measurements. NILM that is full of interest for smart grid applications, allows a better control and understanding of the electricity consumption through a possible real-time feedback. In the literature, most of the state-of-the-art methods disaggregate the observed energy curve measured at the building electrical panel by computing signal features that allow the recognition of the HEA signature [10–12]. Previous works aim at computing an effective set of electrical features that are used in a machine learning framework [13–15]. A recent study [16] proposes a deep learning approach based on CNN applied on a so called "VI trajectory" for which the authors report an average F-measure of 77.60% on the publicly available PLAID dataset [17].

Thus, investigating HEA identification allows us to comparatively assess in a practical scenario the use of several STFT-based TF representations with different TF concentration and to establish links with our previous works [15,18,19] based on "handcrafted" physically meaningful electrical features computed from the active and reactive powers. Our Time–Frequency Representation (TFR) are investigated with several deep CNN architectures.

This paper is organized as follows. In Section 2, we recall the definitions of the STFT-based reassignment and the synchrosqueezing methods. In Section 3, we address the NILM problem, for which we propose several techniques based on different CNN architectures and different TF representations as inputs. We comparatively assess the investigated methods in Section 4. Section 4.3 describes the Layer-wise Relevance Propagation (LRP) method that is used to inspect the CNN feature selection, and the obtained results are analyzed. Finally, in Section 5, we discuss our approach and results with a consideration for future work.

2. STFT-Based Time–Frequency Analysis

Here, we define the tools allowing us to compute the TF representation of a signal that is considered as the input of a neural network architecture.

2.1. Short-Time Fourier Transform (STFT)

Given a signal x, we define its STFT $F_x^h(t,\omega)$ at any time t (expressed in seconds) and frequency ω (expressed in rad.s^{-1}), using a differentiable analysis window h, as:

$$F_x^h(t,\omega) = \int_{\mathbb{R}} x(u)h(t-u)^* \, \mathbf{e}^{-j\omega u} \, \mathrm{d}u, \tag{1}$$

$$= \mathbf{e}^{-j\omega t} \int_{\mathbb{R}} x(t+u)h(-u)^* \, \mathbf{e}^{-j\omega u} \, \mathrm{d}u \tag{2}$$

with $j^2 = -1$, and z^* being the complex conjugate of z. Thus, a TF representation is provided by the spectrogram that can be computed as $|F_x^h(t,\omega)|^2$.

The STFT of a signal admits the following synthesis formula when $h(t_0) \neq 0$, allowing to recover the signal x with a delay $t_0 \geq 0$:

$$x(t - t_0) = \frac{1}{h(t_0)} \int_{-\infty}^{+\infty} F_x^h(t, \omega)\, \mathbf{e}^{j\omega(t-t_0)}\, \frac{\mathrm{d}\omega}{2\pi}, \tag{3}$$

which remains valid for any differentiable analysis window h. In practice, $t_0 = 0$ is commonly used for any symmetrical analysis window but, we show in [20] that a suitable choice is $t_0 = \arg\max_t h(t)$, that can be obtained by solving $\frac{dh}{dt}(t) = 0$. In this study, we use the Gabor transform that is defined for a Gaussian analysis window expressed as:

$$h(t) = \frac{1}{\sqrt{2\pi}T}\, \mathbf{e}^{-\frac{t^2}{2T^2}}, \tag{4}$$

where parameter T corresponds to the time-spread of the window.

2.2. Reassignment and Synchrosqueezing

Reassignment and synchrosqueezing are two sharpening techniques designed to improve the readability of a STFT [4]. The main difference between these methods is the reconstruction capability of the synchrosqueezing in comparison to the reassignment method which provides non reversible TF representations. The reversibility property can be satisfied by preserving the phase information by moving the transform instead of its squared modulus, and through a trade-off which consists of moving the values along a unique axis direction (frequency). As a result, the synchrosqueezing transform has a slightly poorer concentration in the time–frequency plane (depending on the waveform of the analyzed signal) in comparison to the reassignment method, however, it admits a signal reconstruction formula.

2.3. Reassignment

The STFT-based reassignment method consists in moving the values of the spectrogram according to the map $(t, \omega) \mapsto (\hat{t}_{(t,\omega)}, \hat{\omega}_{(t,\omega)})$, where $\hat{t}_{(t,\omega)} = \Re\left(\tilde{t}_{(t,\omega)}\right)$ and $\hat{\omega}_{(t,\omega)} = \Im\left(\tilde{\omega}_{(t,\omega)}\right)$, are the so-called reassignment operators. The reassignment operators correspond to instantaneous frequency and group-delay estimators computed at any TF point as [21]:

$$\tilde{t}_{(t,\omega)} = t - \frac{F_x^{\mathcal{T}h}(t,\omega)}{F_x^h(t,\omega)}, \tag{5}$$

$$\tilde{\omega}_{(t,\omega)} = j\omega + \frac{F_x^{\mathcal{D}h}(t,\omega)}{F_x^h(t,\omega)}, \tag{6}$$

with $\mathcal{T}h(t) = th(t)$ and $\mathcal{D}h(t) = \frac{dh}{dt}(t)$. Thus, the reassigned spectrogram is obtained by computing:

$$\mathrm{RF}_x(t,\omega) = \iint_{\mathbb{R}^2} |F_x^h(\tau,\Omega)|^2 \delta\left(t - \hat{t}(\tau,\Omega)\right) \delta\left(\omega - \hat{\omega}(\tau,\Omega)\right) \mathrm{d}\tau\mathrm{d}\Omega. \tag{7}$$

2.4. Synchrosqueezing

The synchrosqueezing transform of the STFT can be deduced from the simplified synthesis formula given by Equation (3). Hence, for any $t_0 \geq 0$ such that $h(t_0) \neq 0$, the synchrosqueezed STFT can be defined as [22]:

$$\mathrm{SF}_x^h(t,\omega) = \int_{\mathbb{R}} F_x^h(t,\omega')\, \mathbf{e}^{j\omega'(t-t_0)} \delta(\omega - \hat{\omega}_{(t,\omega')})\, \mathrm{d}\omega'. \tag{8}$$

As a result, we can obtain a sharpened TF representation by computing $|\mathrm{SF}_x^h(t,\omega)|^2$. The most important property of $\mathrm{SF}_x^h(t,\omega)$ is its reconstruction capability which leads to:

$$\hat{x}(t-t_0) = \frac{1}{h(t_0)^*}\int_{\mathbb{R}} \mathrm{SF}_x^h(t,\omega)\frac{\mathrm{d}\omega}{2\pi} \tag{9}$$

and allows us to recover the whole signal x. The signal components can also be individually extracted as proposed in [23] when the integration area of Equation (9) is restricted to the vicinity of each ridge.

2.5. Time-Reassigned Synchrosqueezing

In 2019, He et al. proposed, in [24], a new variant of the synchrosqueezing technique which overcomes the problem of non reversibility by moving the transform values along the time axis instead of the frequency axis. Hence, the time-reassigned synchrosqueezed STFT (also called first-order horizontal synchrosqueezing) can be defined as [25]:

$$S_x^h(t,\omega) = \int_{\mathbb{R}} F_x^h(\tau,\omega)\delta\left(t-\hat{t}_x(\tau,\omega)\right)\,d\tau \tag{10}$$

where $\hat{t}_x(t,\omega)$ corresponds to the time reassignment operator which is classically computed using Equation (5). The marginalization over time of the resulting transform leads to:

$$\int_{\mathbb{R}} S_x^h(t,\omega)dt = \iint_{\mathbb{R}^2} F_x^h(\tau,\omega)\delta\left(t-\hat{t}_x(\tau,\omega)\right)\,dtd\tau \tag{11}$$

$$= \int_{\mathbb{R}} F_x^h(\tau,\omega)d\tau = F_h(0)^* F_x(\omega). \tag{12}$$

Hence, an exact signal reconstruction can be obtained using the inverse Fourier transform as [25]:

$$x(t) = \frac{1}{2\pi F_h(0)^*}\iint_{\mathbb{R}^2} S_x^h(\tau,\omega)\,\mathbf{e}^{j\omega t}\,d\tau d\omega. \tag{13}$$

2.6. Discretization

In real-world application scenarios such as NILM, continuous signals are discretized according to a given sampling rate F_s (expressed in Hz). Thus, all the previously defined expressions can be reworded using the rectangle approximation method as $F_x^h[n,m] \approx F_x^h(\frac{k}{F_s}, 2\pi\frac{mF_s}{M})$, where $k \in \mathbb{Z}$ is the time sample index and $m \in [0, M-1]$ is the discrete frequency bin where the number of frequency bins M defines the frequency resolution. As a result, the discrete-time STFT is computed as:

$$F_x^h[n,m] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k]^* e^{-j\frac{2\pi km}{M}} \tag{14}$$

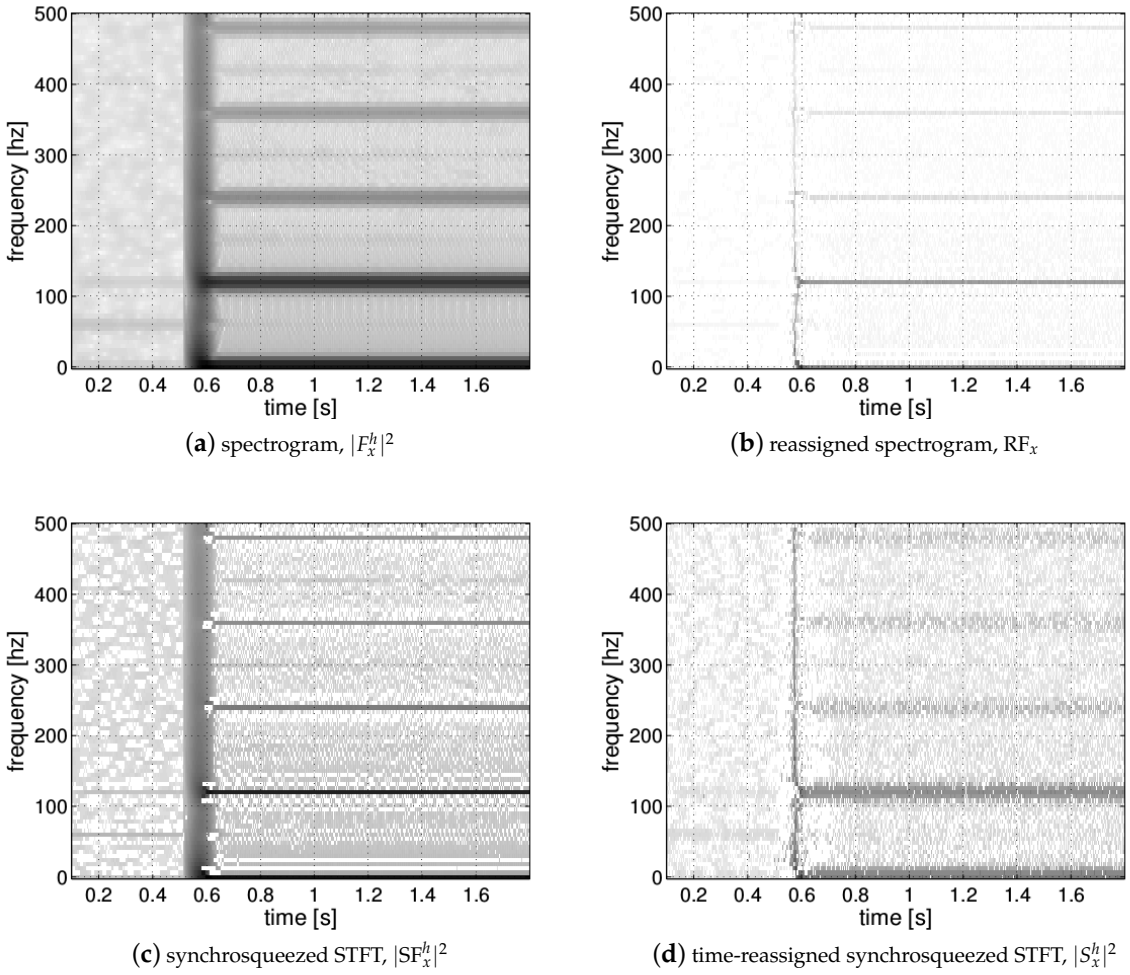
$$= e^{-j2\pi\frac{nm}{M}}\sum_{k=-\infty}^{+\infty} x[k+n]h[-k]^* e^{-j\frac{2\pi km}{M}} \tag{15}$$

For a finite length signal, with a window h defined on $[-K/2, K/2]$, the bounds of the summation can be replaced by $K_{min} = \max(0, n-K/2)$ et $K_{max} = \min(M-1, n+K/2-1)$. Hence, Equation (15) can be computed by:

$$F_x^h[n,m] = e^{-j2\pi\frac{nm}{M}}\sum_{k=K_{min}}^{K_{max}} x[k+n]h[-k]^* e^{-j\frac{2\pi km}{M}} \tag{16}$$

$$= e^{-j2\pi\frac{m}{M}(n+K_{min})}\sum_{k=0}^{K_{max}-K_{min}} x[n+k+K_{min}]h[-(k+K_{min})]^* e^{-j\frac{2\pi km}{M}} \tag{17}$$

As the Gaussian analysis window has an infinite-support length, a threshold Γ is used to obtain a finite number of time samples K_m to be considered for the integration at frequency bin m such as $h[k] \geq \Gamma, \forall\, k \in [-K_m, K_m]$. Hence, a small value for Γ leads to a large value for K_m and an increase of the computation time. Our implementation based on the Gabor transform uses the discrete-time version of the Gaussian window expressed as: $h[n] = \frac{F_s}{\sqrt{2\pi L}}\, \mathbf{e}^{-\frac{n^2}{2L^2}}$ with $L = TF_s$ and with a threshold $\Gamma = 10^{-4}$ which provides good results with a reasonable computation time in practice. A Comparison of the different TFRs computed using the proposed methods applied to a real-world instantaneous power electrical signal is presented in Figure 1.



(**a**) spectrogram, $|F_x^h|^2$ (**b**) reassigned spectrogram, RF_x

(**c**) synchrosqueezed STFT, $|\mathrm{SF}_x^h|^2$ (**d**) time-reassigned synchrosqueezed STFT, $|S_x^h|^2$

Figure 1. Comparison of the different Time–Frequency Representations (TFRs) respectively provided by the Short-time Fourier Transform (STFT): (**a**) the reassigned spectrogram, (**b**) the synchrosqueezing, (**c**) and the time-reassigned synchrosqueezing (**d**) for the instantaneous power signal of a measured appliance from the PLAID dataset [17].

3. Non-Intrusive Load Monitoring

3.1. Problem Formulation

We address the NILM problem through a supervised machine learning approach illustrated in Figure 2, which requires an annotated dataset of individually recorded HEA for the training step. In our approach, the voltage and current real-valued signals respectively denoted $v(t)$ and $i(t)$ correspond to the input of the proposed method. In a realistic application scenario, an electricity consumption measurement system such as the one we proposed in [18] can be used. The training step

(cf. Figure 2a) consists of minimizing the loss function $\mathcal{L}(\hat{y}, y)$ computed between the predicted label $\hat{y}$ and the ground truth y.

The input signals are then used to compute a TF representation from which we derive a set of features used to model each labeled HEA with a unique signature during the training step. Finally, the trained model is used for HEA recognition using the same set of computed input features. The resulting classification method allows the recognition of the closest pre-trained HEA referred to as a "class" according to the machine learning classification terminology.

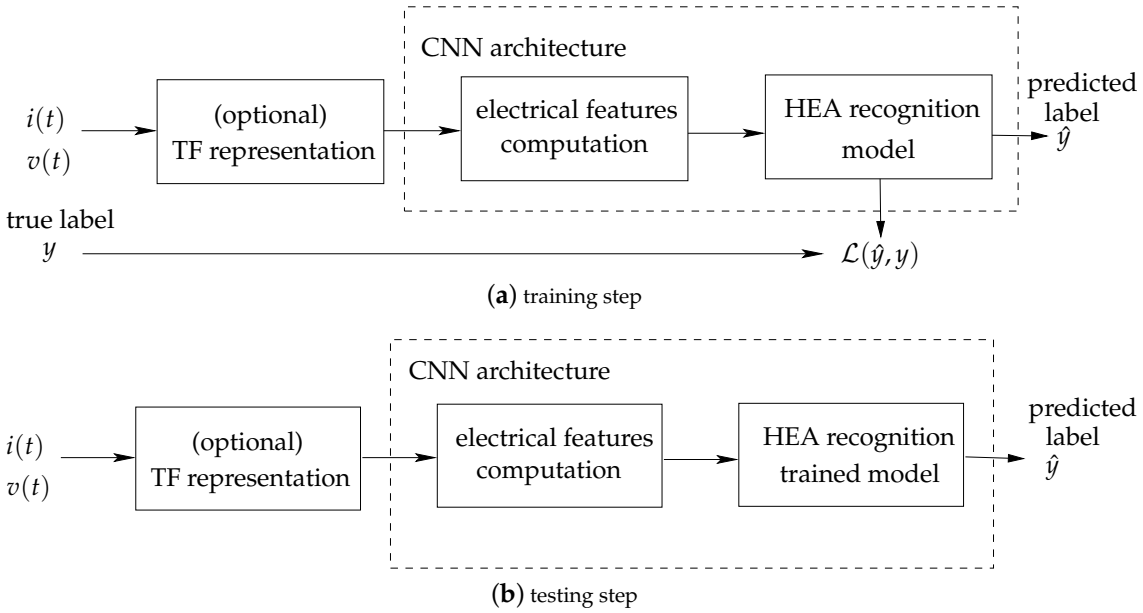


Figure 2. Illustration of the general process for supervised Home Electrical Appliance (HEA) recognition from voltage and current measurements.

3.2. *Electrical Features Computed from Current and Voltage Measurements*

In our previous work [15,19], we introduced a set of 90 features, summarized in Figure 3, which are based on the latest IEEE 1459-2010 standard for the definition of single phase physical components under non-sinusoidal conditions [26,27]. Among the 90 proposed features, we selected 34 related to the active and the reactive power (respectively denoted P and Q) which satisfy the additivity criterion defined in [28]. This property is required for allowing a disaggregation of the mixture signal resulting from several HEAs simultaneously activated before performing the individual HEA recognition process. To this end, an event detection method such as [19] can be used to separate the distinct contribution when several HEAs are switched on simultaneously. In [15], the proposed features are computed for each voltage period and are derived from the Fourier coefficients corresponding to the 15 first harmonics of the voltage frequency. In the present study, we extend the feature definition using the STFT of the discrete-time sampled signals approximated by: $v[n] = v(\frac{n}{F_s})$ and $i[n] = i(\frac{n}{F_s})$, as detailed below.

$p = 90$ electrical features
• Effective current I and its harmonics I_k for $k \in \{1, \ldots, 15\}$ and I_H (A)
• Active power P and its harmonics P_k for $k \in \{1, \ldots, 15\}$ and P_H (W)
• Reactive power Q and its harmonics Q_k for $k \in \{1, \ldots, 15\}$ and Q_H (VAR)
• Apparent power S and its harmonics S_k for $k \in \{1, \ldots, 15\}$ and S_H, S_N (VA)
• Current harmonic distorsion $\mathrm{THD_I}$
• Distorsion power D, D_I, D_V (VAD)
• Power factor F_P, and its harmonics F_{P_k}, for $k \in \{1, \ldots, 15\}$
• Current crest factor F_{C_I}

Figure 3. Description of hand-crafted electrical features proposed in [15].

3.2.1. Electrical Features Based on Fourier Coefficients

For any frequency index (also called bin) $m \in [0, M-1]$, with $M = N$ (N being the finite length of the input signal x), the discrete-time Fourier transform of a signal x can formulated as:

$$F_x[m] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi nm}{M}} \tag{18}$$

$$= \sum_{n=0}^{N-1} x[n] \left(\cos(2\pi n \frac{m}{M}) - j \sin(2\pi n \frac{m}{M}) \right) \tag{19}$$

$$= \sum_{n=0}^{N-1} x[n] \cos(2\pi n \frac{m}{M}) - j \sum_{n=0}^{N-1} x[n] \sin(2\pi n \frac{m}{M}) \tag{20}$$

$$= \frac{N}{2} (x_a[m] - j x_b[m]) \tag{21}$$

where $x_a[m]$ and $x_b[m]$ stand for the Fourier coefficients. As this transform is M-periodic, the angular frequency expressed in rad.s^{-1} associated to a bin m is:

$$\omega = \begin{cases} 2\pi \frac{m}{M} F_s & \text{if } m \in [0, M/2] \\ 2\pi (\frac{m-M}{M}) F_s & \text{if } m \in [M/2+1, M-1] \end{cases} \tag{22}$$

where $m = M/2$ is the integer corresponding to the Nyquist frequency, according to the Shannon sampling theorem. In practice, M can be artificially increased through zero-padding (i.e., addition of 0 values at the end of signal). This process improves the readability but not the frequency resolution of the output which depends of the analysis window properties [2]. With this in mind, we can express the following quantity, where P and Q respectively stand for the active power and reactive power:

$$\frac{4}{N^2} F_v[m] F_i[m]^* = (v_a[m] - j v_b[m])(i_a[m] + j i_b[m]) \tag{23}$$

$$= v_a[m] i_a[m] + j v_a[m] i_b[m] - j v_b[m] i_a[m] + v_b[m] i_b[m] \tag{24}$$

$$= (v_a[m] i_a[m] + v_b[m] i_b[m]) + j(v_a[m] i_b[m] - v_b[m] i_a[m]) \tag{25}$$

$$= P[m] + jQ[m]. \tag{26}$$

3.2.2. New Proposed STFT-Based Electrical Features

The STFT can be viewed as a windowed variant of the Fourier transform, suitable for dealing with non-stationary signals. The main interest of this approach is the capability to convey information about time-varying signal components and the ability to capture transients and impulsive signal components.

Following this idea, we can now replace in Equation (26) the discrete Fourier transforms of the voltage v and the current i by their discrete STFT as defined by Equation (16). Omitting the scaling factor $\frac{4}{N^2}$, we can introduce the following quantity related to the active power P and the reactive power Q:

$$F_v^h[n,m]F_i^h[n,m]^* = P^h[n,m] + jQ^h[n,m] \tag{27}$$

where $F_v^h[n,m]$ and $F_i^h[n,m]$ respectively stand for the STFT of the voltage and the current signals using an analysis window h. This TF representation can thus be used as the input of a CNN architecture to predict the label of an observed HEA.

3.3. Proposed CNN Architectures

The success of CNN is mostly due to their success in image classification tasks and to the development of graphical processing units (GPU), which are of interest for intensive computing. They have successfully been used in a wide variety of signal processing tasks such as speech recognition [29] or music analysis [30], where they are combined with TF representations.

Here, we propose to apply a CNN to the task of HEAs recognition under the assumption that a TF representation conveys a sufficient amount of information valuable in learning useful features. Despite electrical features computed from current and voltage signals frequently being designed and analyzed for HEAs' identification [9,31,32], we expect that the CNN is able to autonomously learn new features based on relevant patterns present in the data. In our study, we compare two distinct inputs which require dedicated CNN architectures:

1. We compute the TF representation of the instantaneous power signal defined as:

$$s[n] = v[n]i[n]. \tag{28}$$

   The spectrogram of this signal looses information about the active and reactive powers. However, it has the advantage of producing a single real-valued matrix that can easily be processed by a classical single input CNN architecture.
2. We compute the product between the voltage TF representation and the complex conjugate of the current TF representation according to Equation (27) which produces a complex matrix $X = P + jQ$. The resulting two-dimensional tensor that contains the real and the imaginary parts, can be processed with the proposed CNN architectures. Our first CNN architecture uses a two-channel model and the second one separately process the real and the imaginary part through two distinct CNNs for which their outputs are concatenated at the last layer.

In what follows, we describe the different types of CNN architectures which attempt to exploit all the available information provided by the distinct computed signal representations.

3.3.1. Single-Input CNN Architecture

The CNN architecture depicted in Figure 4 uses as input the TF representation computed from the instantaneous power signal given by Equation (28). This input layer is connected to four blocks of 32 convolutional neurons with a 3×3 kernel and a REctified Linear Unit (ReLU) activation function followed by a dropout layer and a max-pooling layer. The last hidden convolutional block is connected to two hidden fully connected layers followed by a softmax activation function which produces the final output with the same dimension as the number of expected classes (61 classes considered in this study). Each output node corresponds to the confidence score obtained for each input data. The classification decision rule consists in selecting the HEA class label which obtains the maximal confidence score.

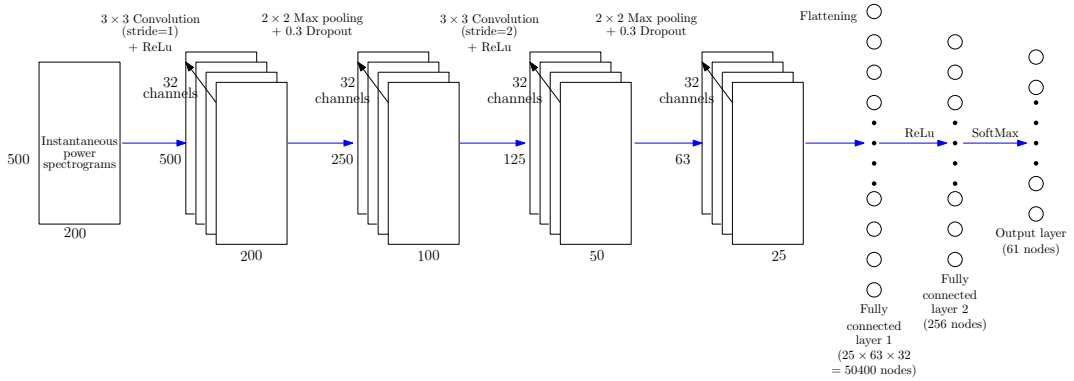


Figure 4. Proposed Convolutional Neural Network (CNN) architecture used for predicting the label of a HEA from the time-frequency representation of its instantaneous power signal.

3.3.2. Two-Channel Input CNN Architecture

To bind the information conveyed by the active and reactive powers, the real and imaginary parts of matrix X that is computed using Equation (27) are processed as separated channels into the same CNN model. Both 2D matrices are merged in one tensor of dimension $M \times N \times 2$ where P and Q are considered in a depth channel as proposed in [33]. Thus, the proposed architecture in Figure 5 is almost identical to the Single-input CNN architecture in Figure 4, except that each convolution kernel contains an additional dimension, resulting in two channels.

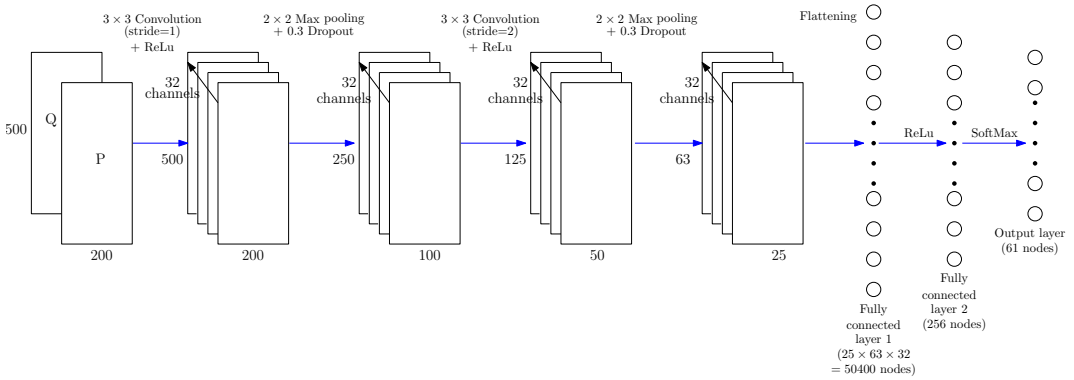


Figure 5. Two-channel input CNN architecture used for predicting the label of a HEA from the time-frequency representations of its active power (P) and reactive power (Q) signals.

3.3.3. Concatenated CNN Architecture

We propose an alternative CNN architecture depicted in Figure 6. It consists of a combination of two convolutional neural networks for which their final outputs are concatenated. The TF representations of the active and reactive powers are simultaneous processed by two independent single-channel CNNs (depicted in Figure 4). Thus, this model has two branches similar to the single-input CNN architecture. A concatenation layer is used to merge feature maps that are processed by two fully connected hidden layers to produce a single final output.

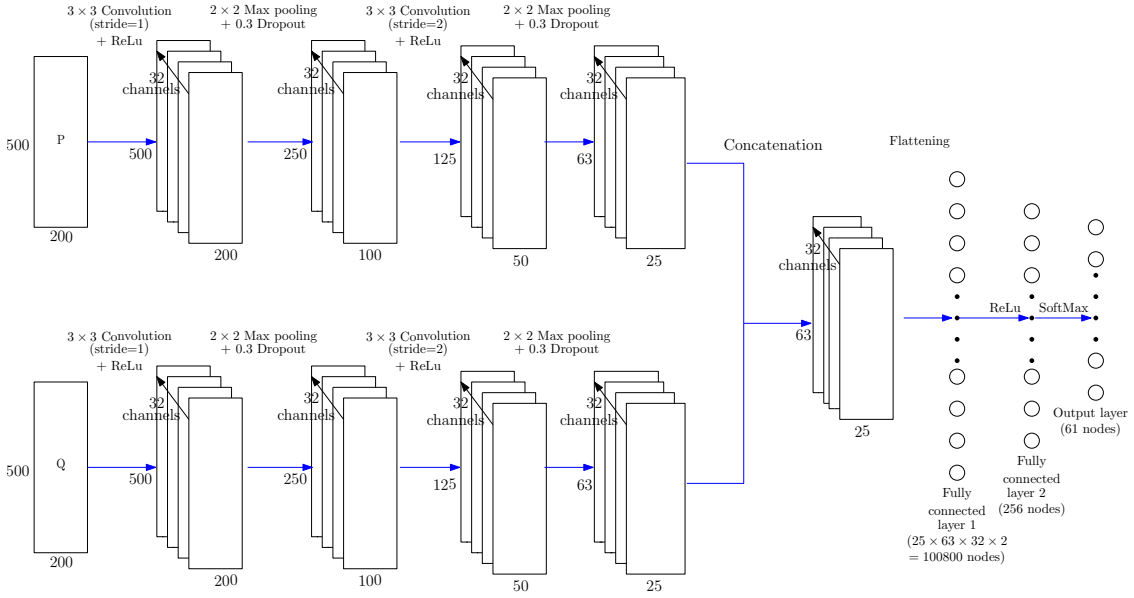


Figure 6. Concatenated CNN architecture used in this research work. The input corresponds to one channel of active power (P) TF representation and one channel of reactive power (Q) TF representation.

4. Numerical Results

4.1. Materials

To make our experiments reproducible, we investigate the publicly available Plug Load Appliance Identification Dataset (PLAID) dataset [17] which contains current and voltage measurements in an American power network at a frequency of 60 Hz. This dataset corresponds to long recordings of start-up transients and steady-state operations of 11 different types of HEA: air conditioner, compact fluorescent lamp, fridge, hairdryer, laptop, microwave, washing machine, bulb, vacuum, fan, heater. Each type of HEA is represented by more than ten different instances. For each HEA, three to six measurements are collected for each state transition. The current and voltage measurements of each HEA were post-processed to extract several windows of 800 samples of the current and voltage waveforms such that we have a total of 61 classes of HEA and 16,476 distinct recordings corresponding to steady-state operations or start-up transients.

The resulting TF representations were computed on recordings of 800 samples using a frequency resolution of $M = 1000$ bins. Since the current and voltage signals are real valued, only the positive frequency bins were considered. We then obtain TF representations of size (500×800). Due to memory limitations, we split each TF representation [34,35] into four parts of size (500×200), which are processed as independent inputs and annotated with the same label as the HEA from which the spectrograms are generated. As a result, our experiment considers a total of $n = 16{,}476 \times 4 = 65{,}904$ individuals.

4.2. HEA Recognition Results

We address the NILM problem as an electrical load recognition experiment where the reference PLAID dataset is randomly partitioned to obtain the following split: training set (80%) and testing set (20%). Then, we compare the results obtained with the different possible TFRs which are either the spectrogram, the synchrosqueezed STFT or the reassigned spectrogram combined with the different CNN architectures. We also compare these results with our previously proposed approach based on hand-crafted electrical features [15,19]. For each considered experiment, exactly the same training and testing samples were used to comparatively assess the methods. The TFRs were computed using our previously proposed MATLAB implementation included into the ASTRES toolbox [22] with $M = 1000$ and Gaussian window with $L \in \{60, 600\}$. The neural networks and the machine

learning experiments were implemented in Python using Keras and Tensorflow frameworks [36]. During the training step, the number of epochs is set to 40 and an early stopping rule is defined to stop the training when the model accuracy converges (considering 20 epochs for the trigger). The training computation time for each CNN model reaches a maximum of 3 h using a NVIDIA GTX 1080 GPU with 8 GB of RAM.

Table 1 shows the comparative classification results of the proposed CNN models combined with each TF representation. These results are expressed in terms of the classification metrics: F-measure, accuracy, recall and precision that can easily be computed from the resulting confusion matrices [37]. According to Table 1, the overall best results (F-measure of 97.5%) are obtained using the spectrograms of the active and reactive powers combined with the two-channel CNN architecture and are comparable to our method based on P, Q features combined with a Random Forest classification method proposed in [19]. Interestingly, in the single input case based on the TF representation of the instantaneous power, the synchrosqueezing transform obtains the best results (F-measure of 92.1%) and outperforms the spectrogram and the reassignment method. However, when a tensor representation is used for P and Q, the reassignment method outperforms the synchrosqueezing transform despite obtains poorer HEA classification results than those obtained using a classical spectrogram or using our baseline method proposed in [15,19].

More interestingly, we notice that, as expected, the combination of the two signals P and Q lead to the best results since they convey the most amount of information about the HEA signature according to the physical models. However, and despite a better readability of the TF representations provided by the synchrosqueezing and the reassignment methods, these methods do not obtain the best classification results. Thus, one can assume that these post-processing methods, which provide sharpened and sparse representations, have a destructive effect on the relevant information embedded in their TF representations as clearly shown by the HEA recognition scores which are about 10% lower.

Table 1. Comparative results (in percentage) of the different HEA recognition methods applied to the PLAID dataset. The window parameter L is empirically chosen to provide the best results.

	Acc	**F_M**	**Rec**	**Pre**
P, Q + Random Forest [15,19]	97.8	97.7	97.6	97.9
STFT (L = 60, single-input CNN)	87.1	87.2	87.3	88.4
STFT (L = 600, CNN with two channels)	97.7	97.5	97.5	97.9
STFT (L = 600, CNN concatenated)	95.6	95.7	95.5	96.1
Synchrosqueezing (L = 600, single-input CNN)	91.9	92.1	92.4	93.1
Synchrosqueezing (L = 60, CNN with two channels)	85.4	85.0	85.4	86.1
Synchrosqueezing (L = 60, CNN concatenated)	87.2	87.3	87.4	87.9
Time-reassigned synchrosqueezing (L = 60, single-input CNN)	85.8	86.1	86.4	85.9
Time-reassigned synchrosqueezing (L = 60, CNN with two channels)	91.4	91.2	90.9	92.1
Time-reassigned synchrosqueezing (L = 60, CNN concatenated)	92.3	92.3	92.4	91.9
Reassigned spectrogram (L = 600, single-input CNN)	74.4	75.0	74.1	77.3

4.3. Relevance Analysis of the Learned Features

4.3.1. Layer-Wise Relevance Propagation

In order to analyze the HEA patterns learned by the CNN architecture applied on each different TF representation, we use the LRP method [38,39] which allows us to quantify the contributions in the input to compute the classification decision [40,41]. Hence, LRP aims at explaining which input features of a given CNN are the most relevant for obtaining the final classification result. This approach is full of interest, since it is complicated to interpret a CNN model from its learned weight parameters.

To this end, LRP proposes to back-propagate the information from the output to the first layer, to obtain a relevance map of the input.

Given a predicted output, LRP proceeds in a reverse manner by iterating over all layers l of the network and propagating relevance values from neurons of hidden layers step-by-step towards the input. Given two neurons i and j connected at two consecutive layers l and $l+1$, each value R_i corresponds to the contribution of an input x_i that is weighted by w_{ij} when the information is propagated into the neural network for computing to the final prediction. Hence, $z_{ij} = x_i w_{ij}$ models the contribution of neuron i to make the neuron j relevant. This value is then divided by the sum of the contributions at layer l connected to neuron j to enforce the conservation property [42]. In the present work, we use the LRP-0 method that is deduced from the LRP-ϵ definition expressed as [38]:

$$R_i^l = \sum_j \frac{z_{ij}}{\epsilon + \sum_k z_{kj}} R_j^{l+1} \quad \text{with} \quad z_{ij} = x_i^l w_{ij}^{l,l+1} \tag{29}$$

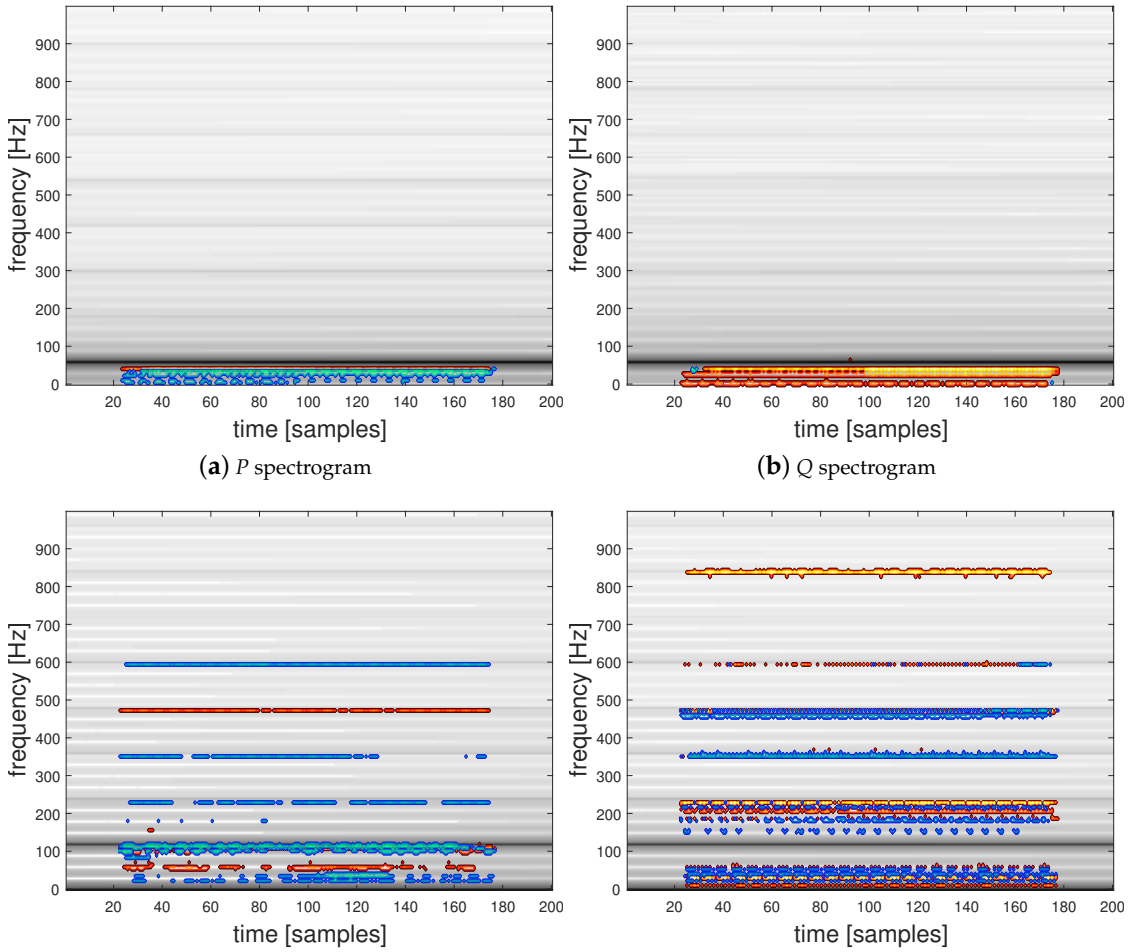
then, the LRP-0 rule is obtained by using $\epsilon = 0$ [42]. These rules satisfy the property such as a null input $x_i = 0$ or the absence of connexion with a weight $w_i = 0$ leads to a null relevance $R_i = 0$. However this rule suffers of a numerical instability when the contributions to the activations of a neuron i are weak or contradictory. It is to be noted that Equation (29) is conservative between layers and satisfies the equality $\sum_i R_i^l = f(x)$ at any layer (l) of the model [43]. In this study, we used the implementation of the algorithm provided by the "iNNvestigate" library [44]. The results are visualized using relevance maps (or heat maps) presented below.

4.3.2. Relevance Maps

We propose to use the LRP-0 method to interpret the discriminatory information contained in the TF representation inputs corresponding to the active and reactive powers used in the two-channel CNN architecture presented in Figure 5. To this end, we present in Figures 7 and 8 spectrograms and the squared modulus of the synchrosqueezed STFT overlayed with the corresponding relevance map computed for two classification examples. Figure 7a,b present the active power P and the reactive power Q spectrograms and Figure 7c,d the squared modulus of synchrosqueezed STFT and the spectrogram of the instantaneous power S with their corresponding overlayed relevance maps respectively computed using the proposed trained 2-channel CNN and the single-input CNN architectures, corresponding to a correctly predicted signal from class 1 (General Electric incandescent light bulb). Figure 8 shows the same TFRs with their corresponding overlayed relevance maps respectively obtained for their trained CNNs, corresponding to a correctly predicted signal from class 32 (a Microwave—Samsung). Interestingly, these TF representations show that, as expected, relevant information is mostly located in the low-frequency range of the signals and mostly embedded around the low-order harmonics. In particular, LRP reveals the area below the first harmonic which contains the most energy located at 120 Hz for S or at 60 Hz for P and Q. The high-frequency area above 700 Hz appears to have a low relevance for each network and lets suppose that high-frequency sampling is of low interest for a HEA recognition application. The computed relevance maps show that there is relevant information between the harmonics of the signal. This is a surprising result in comparison to hand-crafted physics-related features based on the Fourier coefficients of the grid frequency which were previously proposed in [15,19].

To illustrate misclassified examples, we also present in Figure 9, TFRs with their relevance maps for two electrical signals. Figure 9a presents the spectrogram of an individual from class 1 (General Electric incandescent light bulb) which was predicted as class 32 (Samsung microwave) and Figure 9b shows the synchrosqueezed STFT of a signal from class 2 (Frigidaire fridge) predicted as class 16 (Whirlpool washing machine). This latter example clearly present different patterns in their relevance map where the decision was made using the values in high-frequency and between the harmonics parts previously shown as irrelevant. In this case, we believe that the CNN model was trained to use the TFR information located between the harmonics that may be of interest to transient signals (when

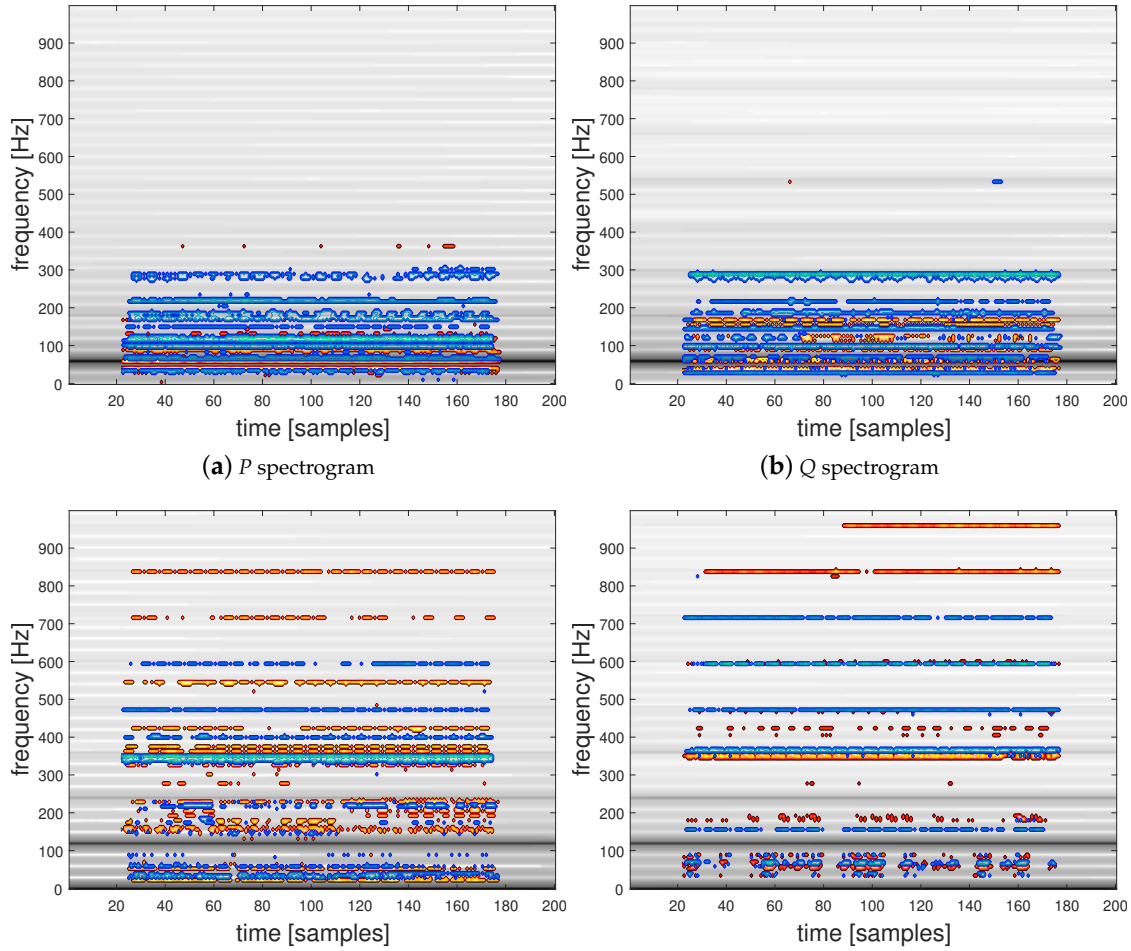
a load is turned on), however, it could result in to a higher sensitivity in the presence of undesired noise which can be explained by electrical network perturbations.



(**a**) *P* spectrogram (**b**) *Q* spectrogram

(**c**) *S* synchrosqueezing (**d**) *S* spectrogram

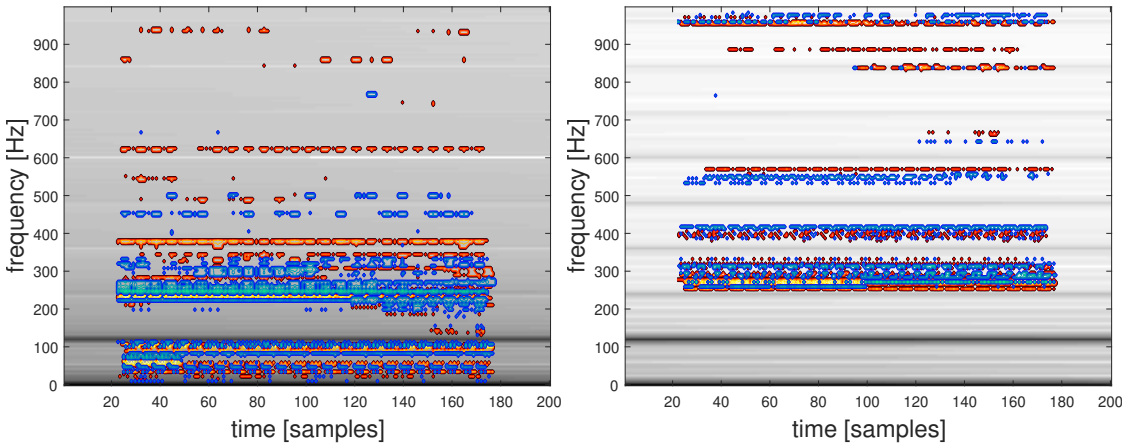
Figure 7. Spectrograms and synchrosqueezed STFT with their corresponding overlayed relevance maps of a correctly predicted individual from class 1 (General Electric incandescent light bulb).



(**a**) *P* spectrogram (**b**) *Q* spectrogram

(**c**) *S* synchrosqueezing (**d**) *S* spectrogram

Figure 8. Spectrograms and synchrosqueezed STFT with their corresponding overlayed relevance maps of a correctly predicted individual from class 32 (Samsung microwave).



(**a**) S synchrosqueezing (**b**) S spectrogram

Figure 9. TFRs with overlayed relevance maps of misclassified individuals. (**a**) corresponds to an individual of class 1 (General Electric incandescent light bulb) predicted as an individual of class 32 (Samsung microwave) and (**b**) corresponds to an individual of class 2 (Frigidaire fridge) predicted as an individual of class 16 (Whirlpool washing machine).

5. Discussion and Future Works

In this paper, we proposed the first study which investigates the use of concentrated TFRs combined with CNNs applied to a realistic HEA recognition task. We performed a comparison between three different forms of TF representations and evaluated their impact on the CNN classification performance of HEAs, using the publicly available PLAID dataset [17]. First of all, the resulting HEA classification scores showed the validity of our approach with excellent results which outperform existing methods of the literature such as [16] which obtained a F-measure of 77.60% on the PLAID dataset. Moreover, our results are comparable to those obtained with the best approach based on hand-crafted electrical features we recently proposed in [15,19].

A further analysis of the trained CNN-based model confirms our physics-related expectations and shows that the usage of active and reactive powers embedded in a tensor representation outperforms the simple usage the apparent power used as a single-input CNN-based architecture. A surprise in our study is that despite good results, the more readable sharpened TF representations provided by the reassignment and synchrosqueezing methods do not provide the best classification results in comparison to classical spectrograms, except using the single-input CNN-based architecture. This result leaves us with the assumption that relevant information is probably lost in the TF representation during the reassignment (resp. synchrosqueezing) process. This may require a further theoretical study based on classical signal models. Finally, investigations based on LRP are made and show that the trained CNN use the information embedded in the low-frequency harmonics and especially the first fundamental harmonic. This result also confirms the relevance of our hand-crafted electrical features proposed in [15,19] and matches to the resulting set of selected features. Thus, this shows that LRP is a suitable solution for explaining which part of the input of a CNN is relevant for reaching a classification decision [45]. As future work, we expect to investigate more complicated CNN architectures such as Recurrent Neural Networks (RNN) and U-net.

Author Contributions: Conceptualization, D.F., S.H. and F.A.; software, D.F., S.H.; writing–original draft preparation, D.F., S.H.; writing–review and editing, F.A.; supervision, D.F..; funding acquisition, D.F., F.A.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the ORACLES Paris-Saclay research project and the French ANR ASCETE project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
HEA	Home Electrical Appliance
LRP	Layer-wise Relevance Propagation
NILM	Non Intrusive Load Monitoring
PLAID	Plug Load Appliance Identification Dataset
ReLU	REctified Linear Unit
STFT	Short-Time Fourier Transform
TF	Time-Frequency

References

1. Flandrin, P. *Explorations in Time-Frequency Analysis*; Cambridge University Press: Cambridge, UK, 2018.
2. Flandrin, P. *Time-Frequency/Time-Scale Analysis*; Academic Press Inc.: Cambridge, MA, USA, 1998.
3. Daubechies, I.; Maes, S. A nonlinear squeezing of the continuous wavelet transform. *Wavelets in Med. and Biol.* **1996**, pp. 527–546. doi:10.1201/9780203734032-20.
4. Auger, F.; Flandrin, P.; Lin, Y.; McLaughlin, S.; Meignen, S.; Oberlin, T.; Wu, H. TF Reassignment and Synchrosqueezing: An Overview. *IEEE Signal Process. Mag.* **2013**, *30*, pp. 32–41.

5. Fourer, D.; Auger, F.; Czarnecki, K.; Meignen, S.; Flandrin, P. Chirp rate and instantaneous frequency estimation: Application to recursive vertical synchrosqueezing. *IEEE Signal Process. Lett.* **2017**, *24*, pp. 1724–1728.
6. Souriau, R.; Fourer, D.; Chen, H.; Lerbet, J.; Maaref, H.; Vigneron, V. High-Voltage Spindles detection from EEG signals using recursive synchrosqueezing transform. In Proceedings of the GRETSI'19, Lille, France, pp. 26–29, August 2019.
7. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
8. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, pp. 1798–1828.
9. Zoha, A.; Gluhak, A.; Imram, M.; Rajasegarar, S. Non intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* **2012**, *12*, pp. 16838–16866.
10. Hart, G.W. Non-intrusive appliance load monitoring. *Proc. IEEE* **1992**, *80*, 1870–1891.
11. Faustine, A.; Mvungi, N.H.; Kaijage, S.; Michael, K. A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem. *arXiv* **2017**, arXiv:1703.00785.
12. Kong, W.; Dong, Z.; Ma, J.; Hill, D.J.; Zhao, J.; Luo, F. An extensible approach for non-intrusive load disaggregation with smart meter data. *IEEE Trans. Smart Grid* **2018**, *9*, pp. 3362–3372.
13. Kato, T.; Cho, H.S.; Lee, D.; Toyomura, T.; Yamazaki, T. Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments. In Proceedings of the International Conference on Smart Homes and Health Telematics, ICOST 2009, Tours, France, 1–3 July 2009; pp. 150–157.
14. Sadeghianpourhamami, N.; Ruyssinck, J.; Deschrijver, D.; Dhaene, T.; Develder, C. Comprehensive feature selection for appliance classification in NILM. *Energy Build.* **2017**, *151*, 98–106.
15. Houidi, S.; Auger, F.; Ben Attia Sethom, H.; Fourer, D.; Miègeville, L. Relevant feature selection for home appliance recognition. In Proceedings of the Electrimacs, Toulouse, France, 4–6 July 2017.
16. De Baets, L.; Ruyssinck, J.; Develder, C.; Dhaene, T.; Deschrijver, D. Appliance classification using VI trajectories and convolutional neural networks. *Energy Build.* **2018**, *158*, 32–36.
17. Gao, J.; Giri, S.; Kara, E.C.; Bergés, M. PLAID: A Public Dataset of High-resolution Electrical Appliance Measurements for Load Identification Research: Demo Abstract. In Proceedings of the ACM Conference on Embedded Systems for Energy-Efficient Buildings, Memphis, TN, USA, 5–6 November 2014; pp. 198–199.
18. Houidi, S.; Auger, F.; Frétaud, P.; Fourer, D.; Miègeville, L.; Sethom, H.B.A. Design of an electricity consumption measurement system for Non Intrusive Load Monitoring. In Proceedings of the 2019 10th International Renewable Energy Congress (IREC), Sousse, Tunisia, 26–28 March 2019.
19. Houidi, S.; Fourer, D.; Auger, F.; Sethom, H.B.A.; Miègeville, L. Home Electrical Appliances Recognition using Relevant Features, Deep learning and Transfer Learning: A Comparative Study. *Energy Build.* **2020**, (submitted).
20. Fourer, D.; Auger, F.; Flandrin, P. Recursive versions of the Levenberg-Marquardt reassigned spectrogram and of the synchrosqueezed STFT. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 4880–4884.
21. Auger, F.; Flandrin, P. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. Signal Process.* **1995**, *43*, pp. 1068–1089.
22. Fourer, D.; Harmouche, J.; Schmitt, J.; Oberlin, T.; Meignen, S.; Auger, F.; Flandrin, P. The ASTRES Toolbox for Mode Extraction of Non-Stationary Multicomponent Signals. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 28 August–2 September 2017; pp. 1170–1174.
23. Meignen, S.; Oberlin, T.; McLaughlin, S. A New Algorithm for Multicomponent Signals Analysis Based on SynchroSqueezing: With an Application to Signal Sampling and Denoising. *IEEE Trans. Signal Process.* **2012**, *60*, pp. 5787–5798.
24. He, D.; Cao, H.; Wang, S.; Chen, X. Time-reassigned synchrosqueezing transform: The algorithm and its applications in mechanical signal processing. *Mech. Syst. Signal Process.* **2019**, *117*, 255–279.
25. Fourer, D.; Auger, F. Second-Order Time-Reassigned Synchrosqueezing Transform: Application to Draupner Wave Analysis. In Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2–6 September 2019.
26. Langella, R.; Testa, A. IEEE standard definitions for the measurement of electric power quantities under sinusoidal, nonsinusoidal, balanced, or unbalanced conditions. In *Revision of IEEE Std. 1459–2000*; IEEE: Piscataway, NJ, USA, 2010; pp. 1–40.

27. Eigeles, E.A. On the Assessment of Harmonic Pollution. *IEEE Trans. Power Deliv.* **1995**, *10*, pp. 693–98.
28. Liang, J.; Ng, S.K.K.; Kendall, G.; Cheng, J.W.M. Load Signature Study Part I: Basic Concept, Structure, and Methodology. *IEEE Trans. Power Deliv.* **2010**, *25*, 551–560.
29. Badshah, A.M.; Ahmad, J.; Rahim, N.; Baik, S.W. Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network. In Proceedings of the 2017 International Conference on Platform Technology and Service (PlatCon), Busan, Korea, 13–15 February 2017; pp. 1–5.
30. Solanki, A.; Pandey, S. Music instrument recognition using deep convolutional neural networks. *Int. J. Inf. Technol. (IJITEE)* **2019**, *8*, pp. 1076–1079.
31. Ruzzelli, A.; Nicolas, C.; Schoofs, A.; O'Hare, G. Real-time recognition and profiling of appliances through a single electricity sensor. In Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Boston, MA, USA, 21–25 June 2010; pp. 1–9.
32. Bouhouras, A.; Gkaidatzis, P.; Paschalis, A.; Chatzisavvas, K.; Panagiotou, E.; Poulakis, N.; Christoforidis, G. Load Signature Formulation for Non-Intrusive Load Monitoring Based on Current Measurements. *Energies* **2017**, *10*, 538.
33. Caracalla, H.; Roebel, A. Sound texture synthesis using RI spectrograms. In Proceedings of the ICASSP 2020—45th International Conference on Acoustics, Speech, and Signal Processing, Barcelona, Spain, 4–8 May 2020; pp. 416–420.
34. Costa, Y.M.G.; de Oliveira, L.E.S.; Silla, C.N. An evaluation of Convolutional Neural Networks for music classification using spectrograms. *Appl. Soft Comput.* **2017**, *52*, pp. 28–38.
35. Phaye, S.S.R.; Benetos, E.; Wang, Y. SubSpectralNet—Using Sub-spectrogram Based Convolutional Neural Networks for Acoustic Scene Classification. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 825–829.
36. Tensorflow-Guide to the Keras Functional API. Available online: https://www.tensorflow.org/overview (accessed on 18 August 2020).
37. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, pp. 37–63.
38. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE* **2015**, *10*, e0130140.
39. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.R. Toward Interpretable Machine Learning: Transparent Deep Neural Networks and Beyond. *arXiv* **2020**, arXiv:2003.07631.
40. Yang, Y.; Tresp, V.; Wunderle, M.; Fasching, P.A. Explaining Therapy Predictions with Layer-Wise Relevance Propagation in Neural Networks. In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI): New York, NY, USA, 4–7 June 2018; pp. 152–162.
41. Lapuschkin, S.; Wäldchen, S.; Binder, A.; Montavon, G.; Samek, W.; Müller, K.R. Unmasking Clever Hans predictors and assessing what machines really learn. *Nat. Commun.* **2019**, *10*, 1096.
42. Montavon, G.; Binder, A.; Lapuschkin, S.; Samek, W.; Müller, K. Layer-Wise Relevance Propagation: An Overview. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer: Berlin, Germany: 2019.
43. Kohlbrenner, M.; Bauer, A.; Nakajima, S.; Binder, A.; Samek, W.; Lapuschkin, S. Towards best practice in explaining neural network decisions with LRP. *arXiv* **2019**, arXiv:1910.09840.
44. Alber, M.; Lapuschkin, S.; Seegerer, P.; Hägele, M.; Schütt, K.T.; Montavon, G.; Samek, W.; Müller, K.R.; Dähne, S.; Kindermans, P.J. iNNvestigate neural networks! *J. Mach. Learn. Res.* **2019**, *20*, pp. 1–8.
45. Binder, A.; Montavon, G.; Lapuschkin, S.; Müller, K.R.; Samek, W. Layer-wise relevance propagation for neural networks with local renormalization layers. In Proceedings of the International Conference on Artificial Neural Networks (ICANN), Barcelona, Spain, 6–9 September 2016; Springer: Berlin, Germany, 2016; pp. 63–71.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).