Scale- and Context-Aware Convolutional Non-intrusive Load Monitoring[⋆]

Kunjin Chen[*], Yu Zhang[†], Qin Wang[#], Jun Hu[*], Hang Fan[*], Jinliang He[*]

[*]Tsinghua University, [†]UC Santa Cruz, [#]ETH Zürich

Funding ACK: Hellman Fellowship, 2019 Seed Fund from CITRIS and the Banatao Institute

[⋆] *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2362–2373, May 2020.

`https://ieeexplore.ieee.org/document/8897031`

What is Energy Disaggregation?

- **Energy disaggregation** (a.k.a. nonintrusive load monitoring, NILM): Decompose the whole energy consumption of a house into the energy usage of individual appliances (single-channel blind source separation problem).

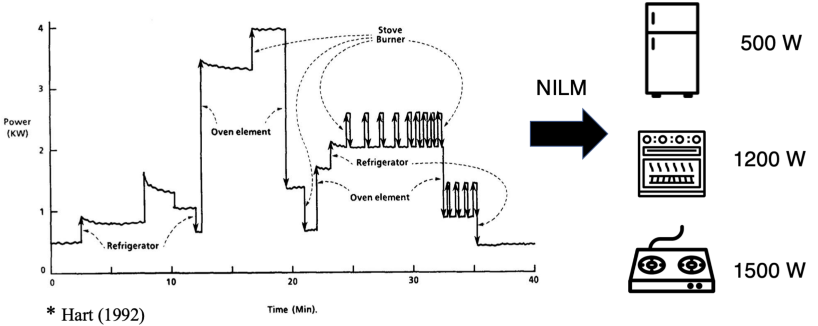
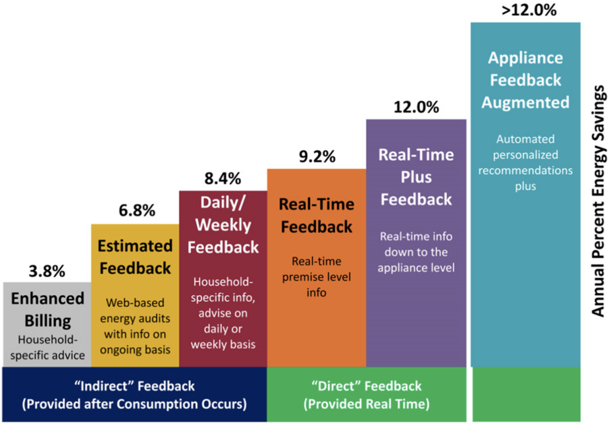


Figure: github.com/ch-shin/awesome-nilm

Why Energy Disaggregation?

- Appliance-level feedback helps reduce energy consumption.



- Benefits for **utilities**: Energy efficiency marketing, incentives and rate structures, program design and evaluation, policy recommendation.

Figure: Armel et al. (2013)

Energy Management via Disaggregation

Companies offer household energy management services via NILM.

ECOISME
eliq
ENCORED
e-sylife be connected to your home
verv
Informetis
sense
plotwatt

bidgely
CHAI ENERGY
ecotagious
Grid4C Foreseeing Power
HOMEpulse SMART ENERGY SERVICES
Intelen ENGAGING PEOPLE TO RETH!NK
ONZO
Qinergy

watt is Intelligent Solutions
smart3
WATTY
smappee
Voltaware
powerley

Source: http://wiki.nilm.eu/companies.html

Energy Management via Disaggregation

An example from Bidgley (15 million households):

Figure: http://pvsolarreport.com/bidgely-solar-disaggregation-platform/

Prior Work

Models and algorithms for energy disaggregation:

- **Supervised**: k-means and SVM [Altrabalsi et al. (2014)], various neural network models including RNN, CNN, and denoising auto-encoders [Kelly and Knottenbelt (2015)].
- **Unsupervised**: factorial hidden Markov model [Kolter and Jaakkola (2012)], graph signal processing [Zhao et al. (2016)].
- **Other methods**: appliance-wise rules based on state changes [Farinaccio and Zmeureanu (1999)].

We propose the scale- and context-aware network (SCANet) in this work.
Incorporate the understanding of the task's challenges into the model:

- Energy consumption behaviors contain higher-level *domain knowledge*: e.g., i) dryer works after washer; and ii) microwave can be turned on multiple times until cooking is finished.
- Time scale variance of different appliances: Requiring the ability to deal with *scale variation*.

Problem Formulation

- Aggregate power consumption signal: $\tilde{\mathbf{x}} = (x_1, \cdots, x_T)$.
- Power consumption of the i-th appliance: $\tilde{\mathbf{y}}^i = (y_1^i, \cdots, y_T^i)$.
- Total power consumption of all remaining appliances: $\tilde{\mathbf{u}} = (u_1, \cdots, u_T)$.

The model:

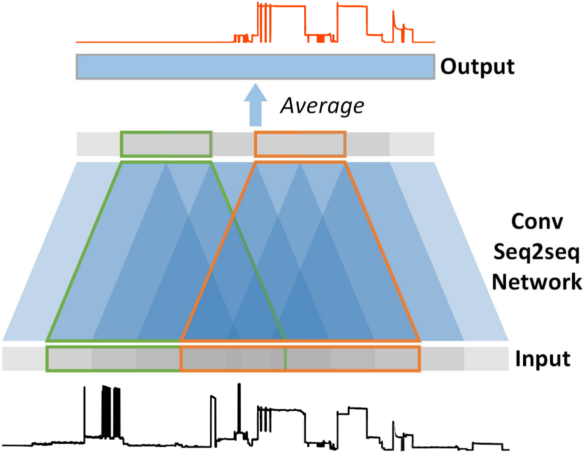
$$\boxed{x_t = \sum_{i=1}^{N_a} y_t^i + u_t + \epsilon_t,}$$

where N_a is the number of appliances and ϵ_t is the additive noise.

The task of energy disaggregation: Given the aggregate signal $\tilde{\mathbf{x}}$, recover the power consumption sequences $\{\tilde{\mathbf{y}}^i\}_{i=1}^{N_a}$ of the appliances under consideration.

Convolutional Neural Networks for NILM

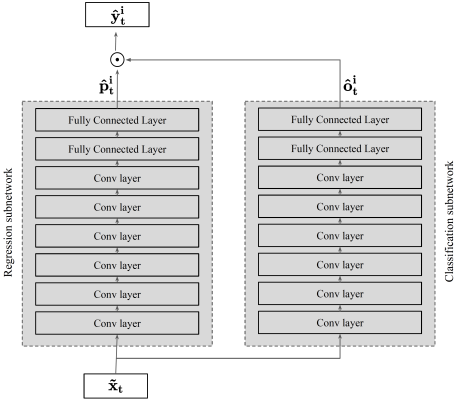
- Use sliding windows to produce input/output samples.
- The input size $(s + 2w)$ is larger than the output size (s) to leverage the contextual information.



Two Joint Learning Tasks

How to independently accomplish the tasks of **regression** & **classification**?

- Estimating the power consumed by appliance i at time t: $\hat{\mathbf{p}}_t^i \in \mathbb{R}_+^s$
- Identifying the on/off state of appliance i at time t: $\hat{\mathbf{o}}_t^i \in [0,1]^s$.



$$\hat{\mathbf{y}}_t^i = \hat{\mathbf{p}}_t^i \odot \hat{\mathbf{o}}_t^i$$

Now, the two sub-networks only need to focus on their own tasks.

Figure credit: Shin et al. (2018)

SCANet vs SGN

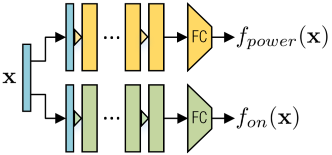


Figure: The structure of the subtask gated networks (SGN) by Shin et al. (2018).

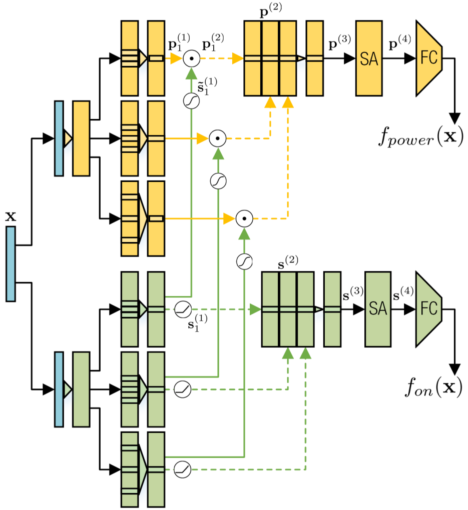


Figure: The structure of the proposed SCANet.

Scale-aware Feature Extraction

Key: Use convolution with different dilation rates (r_d) to generate multi-scale features.

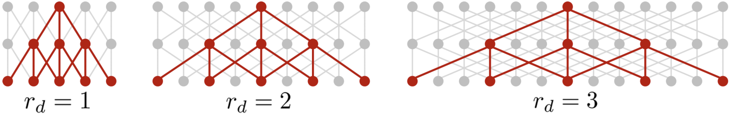


Figure: An illustration of dilated convolution: The space between original kernel elements get wider with the increased dilation rate.

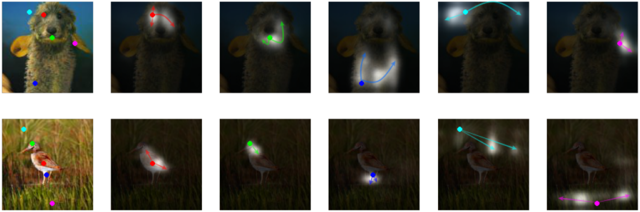
- A larger r_d allows the output nodes respond to wider time ranges at the input. Thus, the outputs at the branches with different r_d will reflect elements (e.g., shapes or edges) of different time scales at the input.
- An element at the input will affect more output nodes for a larger r_d.
- For SCANet, we use 3 branches with $r_d = 1, 2, 3$ in both sub-networks, which are associated as

$$\boxed{\mathbf{p}_j^{(2)} = \mathbf{p}_j^{(1)} \odot \tilde{\mathbf{s}}_j^{(1)},\ j = 1, 2, 3.} \quad (1)$$

Self-attention Module

Self-attention: Imitation of the human sight mechanism.

- Image generation:

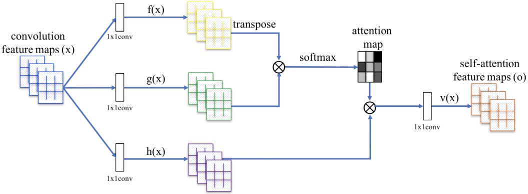


- When the human sight mechanism detects an item, it will typically not scan the entire scene end to end. Rather, it will always focus on a specific portion according to the person's need.
- When we realize that an object often appears in a particular part of a scene, we will learn that for the future and tend to focus our attention on that area.

Figure credit: Zhang et al. (2018), Wang et al. (2018)

Context-aware Feature Integration: Self-attention Module

Self-attention: Values of one time step are obtained by attending to all time steps.



- For input $\mathbf{z} \in \mathbb{R}^{C \times L}$ with L time steps and C channels, we have the mappings: $g(\mathbf{z}) = \mathbf{W}_g \mathbf{z}$ and $h(\mathbf{z}) = \mathbf{W}_h \mathbf{z}$.
- Entry $a_{j,i}$ in the attention matrix $\mathbf{A}$ is calculated as

$$a_{j,i} = \frac{\exp(\tilde{a}_{i,j})}{\sum_{\ell=1}^{L} \exp(\tilde{a}_{\ell,j})}, \text{ where } \tilde{a}_{i,j} = \left[g(\mathbf{z})^\top h(\mathbf{z})\right]_{i,j} \tag{2}$$

- The additional feature map $\mathbf{r}$ is given as

$$\mathbf{r} = d(\mathbf{z})\mathbf{A} = \mathbf{W}_d \mathbf{z} \mathbf{A}. \tag{3}$$

- The module outputs $\mathbf{z} + \gamma \mathbf{r}$.

Figure: Zhang et al. (2018)

Adversarial Loss

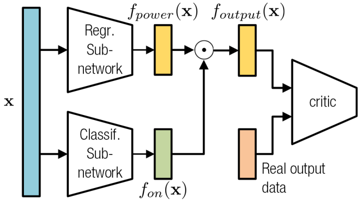
Loss function of the model: $\mathcal{L} = \mathcal{L}_{\text{output}} + \mathcal{L}_{\text{on}}$

- $\mathcal{L}_{\text{output}}$: mean squared error (MSE) for the model output.
- $\mathcal{L}_{\text{on}}$: binary cross-entropy (BCE) for the classification sub-network
  $\mathcal{L}_{\text{on}} = -\frac{1}{N}\sum_{i=1}^{N} o_i \log(\hat{o}_i) + (1 - o_i)\log(1 - \hat{o}_i)$.

Adding an adversarial loss term helps generate **realistic outputs**

$$\mathcal{L} = \mathcal{L}_{\text{output}} + \mathcal{L}_{\text{on}} + \lambda\mathcal{L}_{\text{adv}}.$$

The complete SCANet model with the additional critic module: Formulate the model as a Wasserstein GAN with gradient penalty:

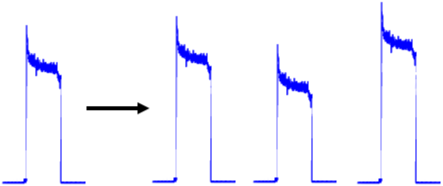


On-state Augmentation

Create data samples with varied power consumption values as data augmentation: e.g., the peak power of two fridge models may be different even if they have the same operation pattern.

- Decide the maximum offset values $e^- < 0$ and $e^+ \geq 0$.
- Replace $\mathbf{x}$ and $\mathbf{y}$ by $\mathbf{x} + e\tilde{\mathbf{o}}$ and $\mathbf{y} + e\mathbf{o}$, where $e \sim \mathrm{unif}(e^-, e^+)$.
- Apply the augmentation during the training of the network (different augmentation for the same $\mathbf{y}$ in each epoch).

An example of on-state augmentation for fridge:



Implementation Details

- **Two real datasets**.
  1) REDD (6 houses in US, every 3 seconds): house 2-6 for training, house 1 for testing.
  2) UK-DALE (5 houses in UK, every 6 seconds): house 1 and house 5 for training, house 2 for testing.
- **Performance metrics**: mean absolute error (MAE); signal aggregate error (SAE, absolute difference of sum over time periods).

- Leverage the CNN backbone in SGN [Shin et al. (2018)] and add the scale- and context-aware modules.
- The input and output sequences cover 43.2 min and 3.2 min, respectively.
- Adversarial loss is added for dishwasher and microwave. On-state augmentation is implemented for fridge (both for REDD dataset).

Estimation Errors

The REDD dataset

Improvement: 22.6% for MAE and 28.2% for SAE

Metric	Model	Fridge	Microwave	Dishwasher	Average
MAE	Seq2point [Zhang *et al.*, 2018a]	26.01	27.13	24.15	25.76
	SGN [Shin *et al.*, 2018]	26.11	16.95	15.88	19.65
	Proposed SCANet	**21.77**	**13.75**	**10.12**	**15.21**
SAE	Seq2point	16.24	18.89	22.98	19.37
	SGN	17.28	12.49	14.59	14.79
	Proposed SCANet	**14.05**	**9.97**	**7.83**	**10.62**

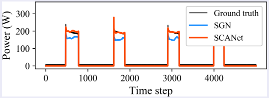
The UK-DALE dataset

Improvement: 9.3% for MAE and 15.5% for SAE

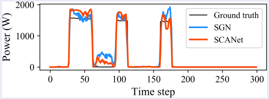
Metric	Model	Washing machine	Kettle	Fridge	Microwave	Dishwasher	Average
MAE	Seq2point	15.09	11.43	18.01	17.87	18.50	16.18
	SGN	10.40	8.22	14.66	6.52	11.11	10.18
	Proposed SCANet	**8.84**	**7.36**	**14.37**	**5.53**	**10.06**	**9.23**
SAE	Seq2point	11.98	5.60	**6.99**	14.62	12.35	10.31
	SGN	7.99	5.35	9.24	5.11	8.43	7.22
	Proposed SCANet	**6.63**	**4.32**	8.50	**4.34**	**6.71**	**6.10**

Results: Examples

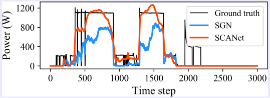
The REDD dataset



(a) Fridge

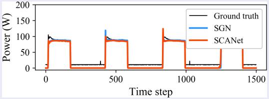


(b) Microwave

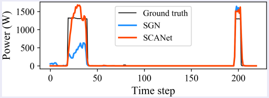


(c) Dishwasher

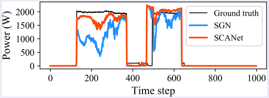
The UK-DALE dataset



(a) Fridge

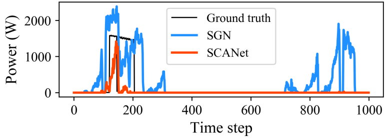


(b) Microwave

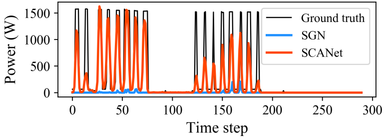


(c) Dishwasher

More Results for Microwave



(a) False positive example of SGN

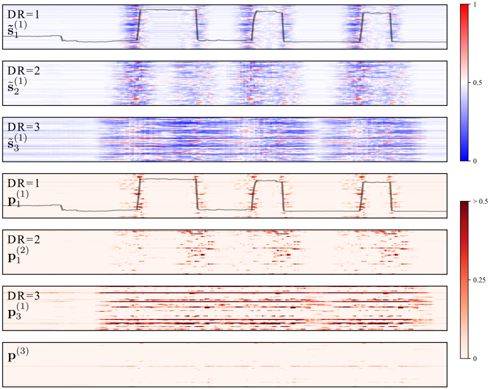


(b) A case SCANet is able to detect

- (a) SGN tends to have more false positive sections.
- (b) SCANet can identify that the microwave consumes power for multiple successive short durations.

Visualization: Multi-scale Features

Visualization of feature maps of multiple scales for microwave in the REDD dataset:



- The multi-scale features are mostly responding to edges.
- The classification subnet gates out lots of features in the regression subnet.

Visualization: Self-attention Mechanism

The self-attention module of the classification sub-network focuses on rising edges of microwave:

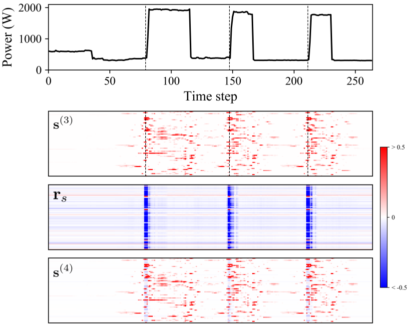


Figure: Feature maps in the classification sub-network for microwave.

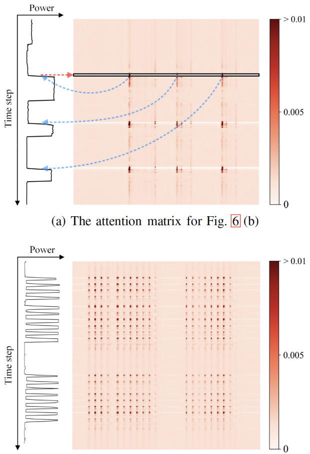
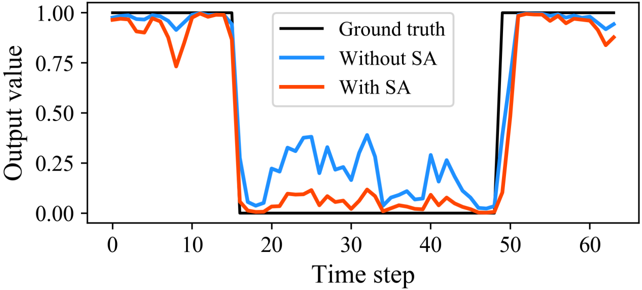


Figure: The self-attention matrix $\mathbf{A}^\top$ in the classification sub-network for microwave.

Visualization: Self-attention Mechanism (Cont'd)

Compare the outputs of the classification sub-network with and without the self-attention module:



Ablation Study

Try different combinations of modules for dishwasher in the REDD dataset (MS: multi-scale, SA: self-attention, AL: adversarial loss)

MS	SA	AL	MAE
-	-	-	15.88
✓			14.33
	✓		13.70
✓	✓		13.15
		✓	13.83
✓	✓	✓	10.12

- The modules are mutually compatible since lower MAEs can be achieved when they are combined.

Takeaway

In this work, we developed a subtask gated convolutional neural network for energy disaggregation. By incorporating the domain knowledge, we proposed a series of performance-enhancing techniques that are listed in below.

1. Improve scale-awareness via dilated convolution.
2. Use an attention mechanism to extract context information.
3. Leverage on-state augmentation to capture heterogeneous power consumption patterns, even for a single type of appliance (e.g., fridge).
4. Facilitate the generation of more realistic outputs via GAN.

Thank You!

References I

Altrabalsi, H., Liao, J., Stankovic, L. and Stankovic, V. (2014). A low-complexity energy disaggregation method: Performance and robustness. In 2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG) pp. 1–8, IEEE.

Farinaccio, L. and Zmeureanu, R. (1999). Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. Energy and Buildings *30*, 245–259.

Kelly, J. and Knottenbelt, W. (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 pp. 55–64, ACM.

Kolter, J. Z. and Jaakkola, T. (2012). Approximate inference in additive factorial hmms with application to energy disaggregation. In Artificial intelligence and statistics pp. 1472–1482,.

Shin, C., Joo, S., Yim, J., Lee, H., Moon, T. and Rhee, W. (2018). Subtask Gated Networks for Non-Intrusive Load Monitoring. arXiv preprint arXiv:1811.06692 .

References II

Wang, X., Girshick, R., Gupta, A. and He, K. (2018). Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 7794–7803,.

Zhang, H., Goodfellow, I., Metaxas, D. and Odena, A. (2018). Self-attention generative adversarial networks. arXiv preprint arXiv:1805.08318 *None*, 1–1.

Zhao, B., Stankovic, L. and Stankovic, V. (2016). On a training-less solution for non-intrusive appliance load monitoring using graph signal processing. IEEE Access *4*, 1784–1799.