TARNet: Task-Aware Reconstruction for Time-Series Transformer

Ranak Roy Chowdhury*, Xiyuan Zhang, Jingbo Shang, Rajesh K. Gupta, Dezhi Hong

University of California San Diego, La Jolla, CA, USA

In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’22)

August 14–18, 2022, Washington, DC, USA

*Primary Author Contact: rrchowdh@eng.ucsd.edu

Code is publicly available at https://github.com/ranakroychowdhury/TARNet

Outline

- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

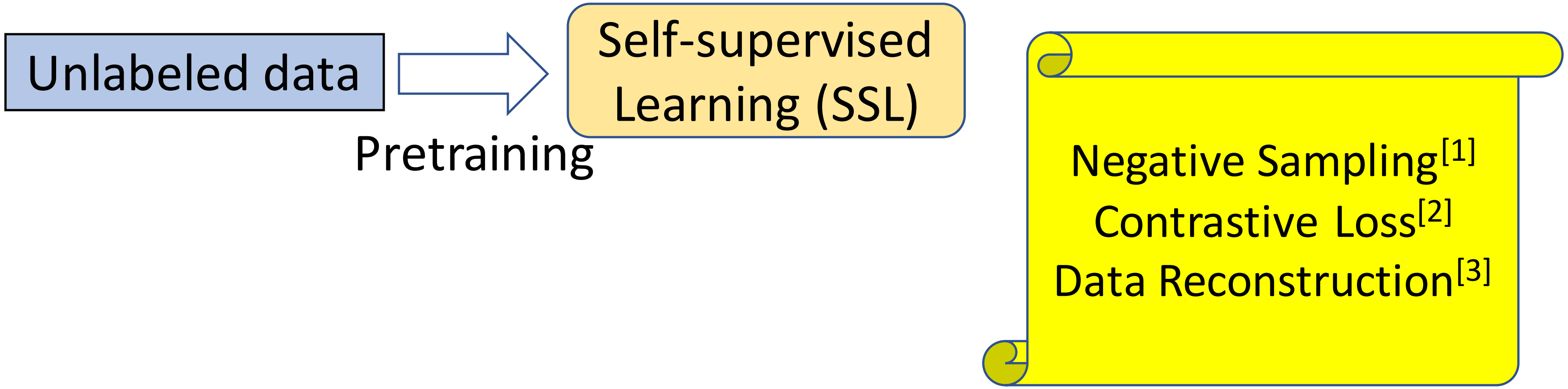
Outline

- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

Research Question

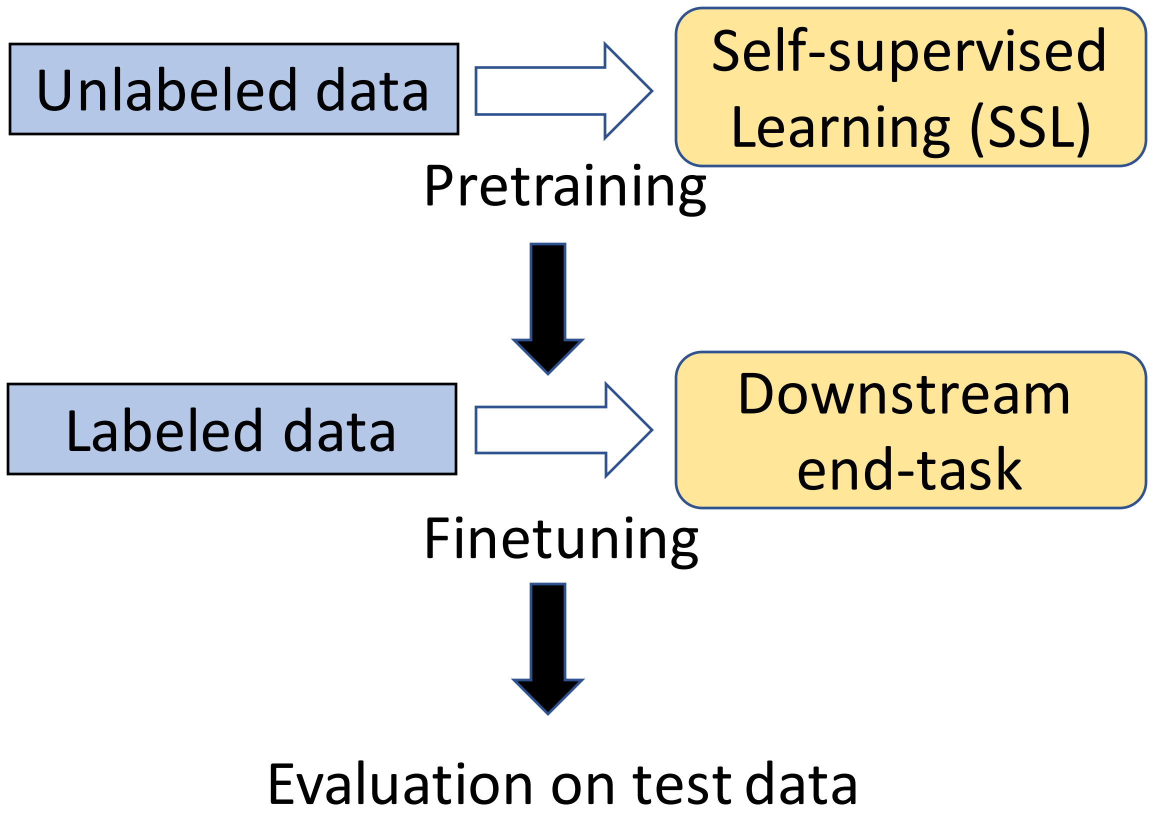
How can we make pre-training for time-series *end-task aware*?

Motivation

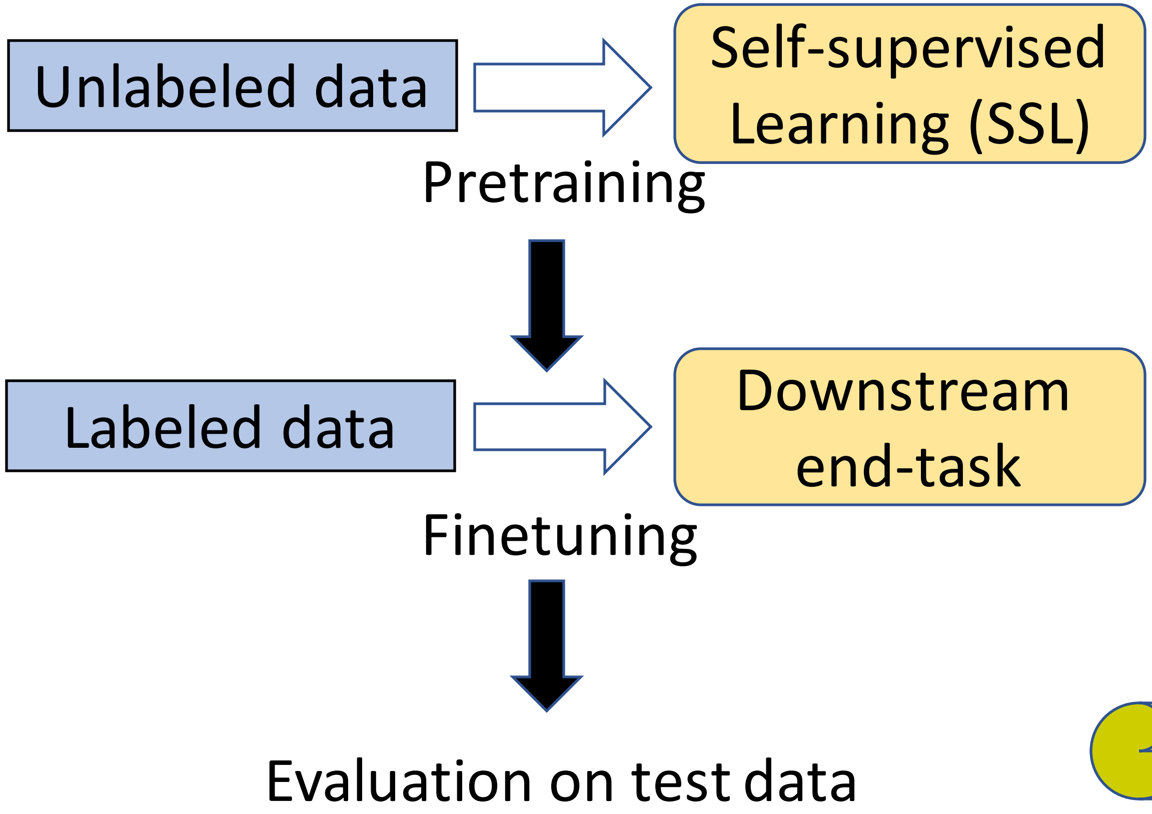


[1] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. 2019. Unsupervised scalable representation learning for multivariate time series. arXiv preprint arXiv:1901.10738 (2019).
[2] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, and Bixiong Xu. 2021. Learning Timestamp-Level Representations for Time Series with Hierarchical Contrastive Loss. arXiv preprint arXiv:2106.10466 (2021).
[3] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. [n. d.]. A Transformer-based Framework for Multivariate Time Series Representation Learning. In KDD, pages=2114–2124, year=2021.

Motivation

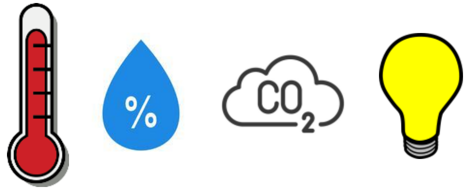

Unlabeled data
Self-supervised Learning (SSL)
Pretraining
Labeled data
Downstream end-task
Finetuning
Evaluation on test data

Motivation



1. SSL is decoupled from end-task
2. Data representation learnt in an *end-task agnostic* manner
3. Sub-optimal performance on test data

Motivation

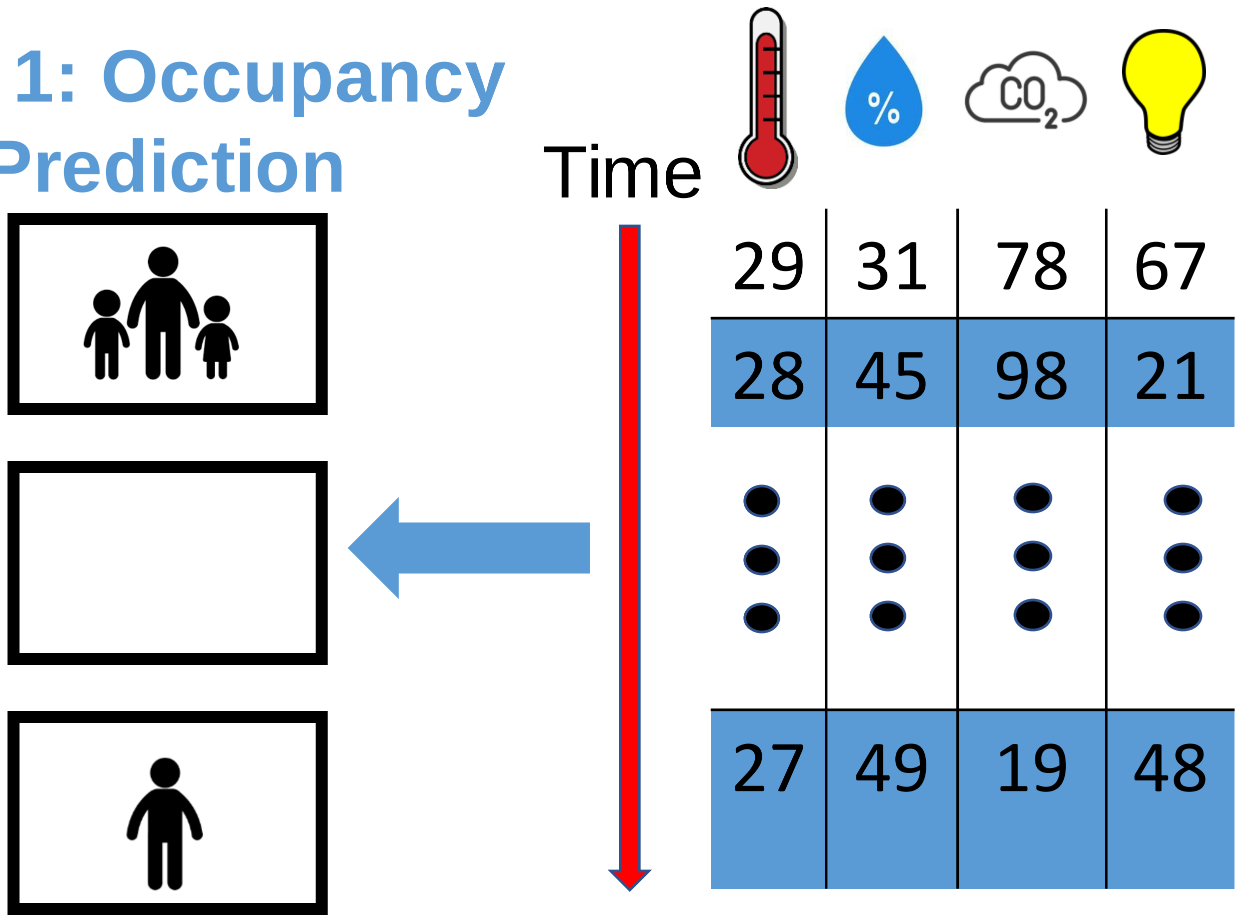


Time

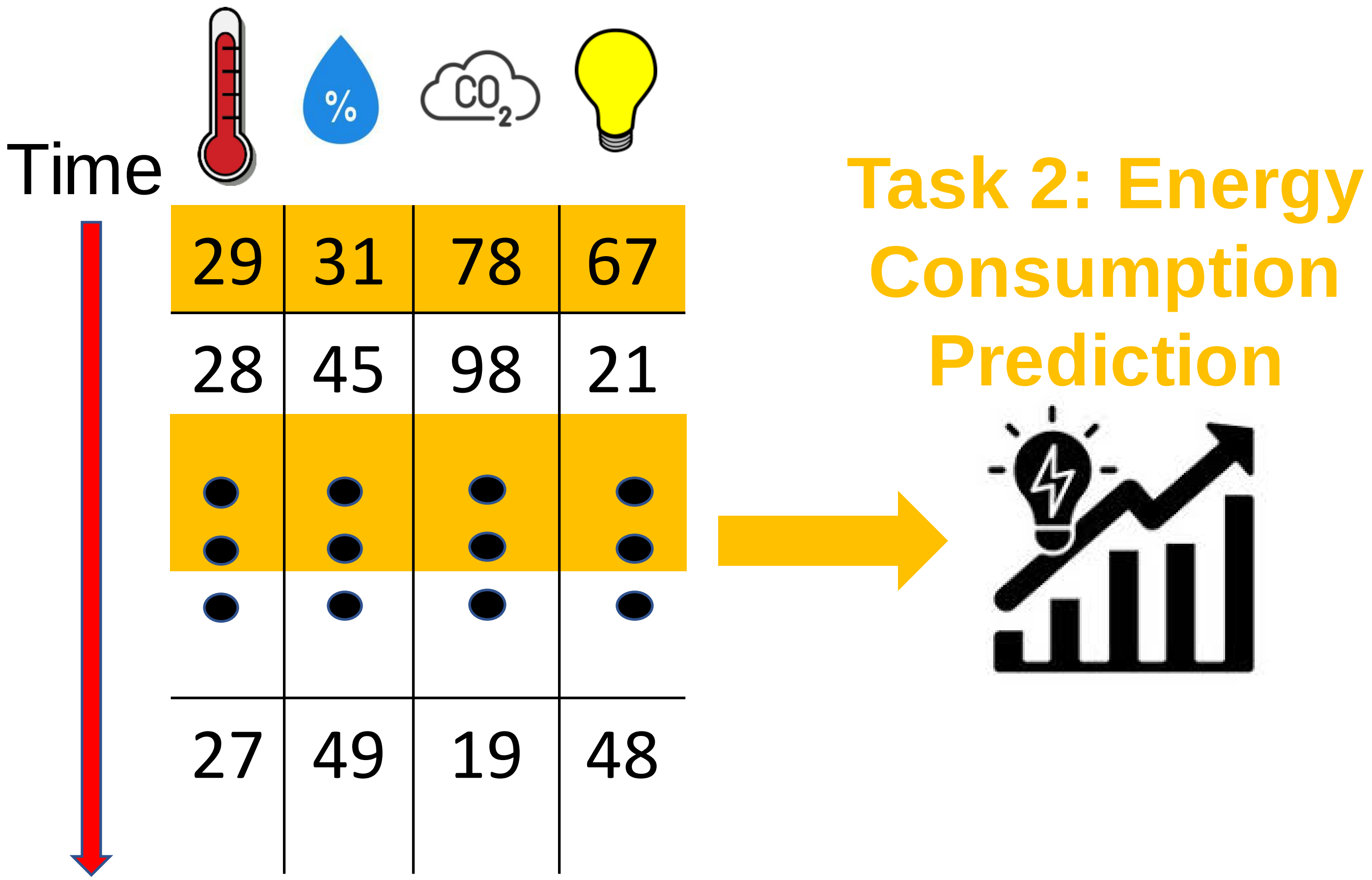
29	31	78	67
28	45	98	21
⋮	⋮	⋮	⋮
27	49	19	48

Motivation

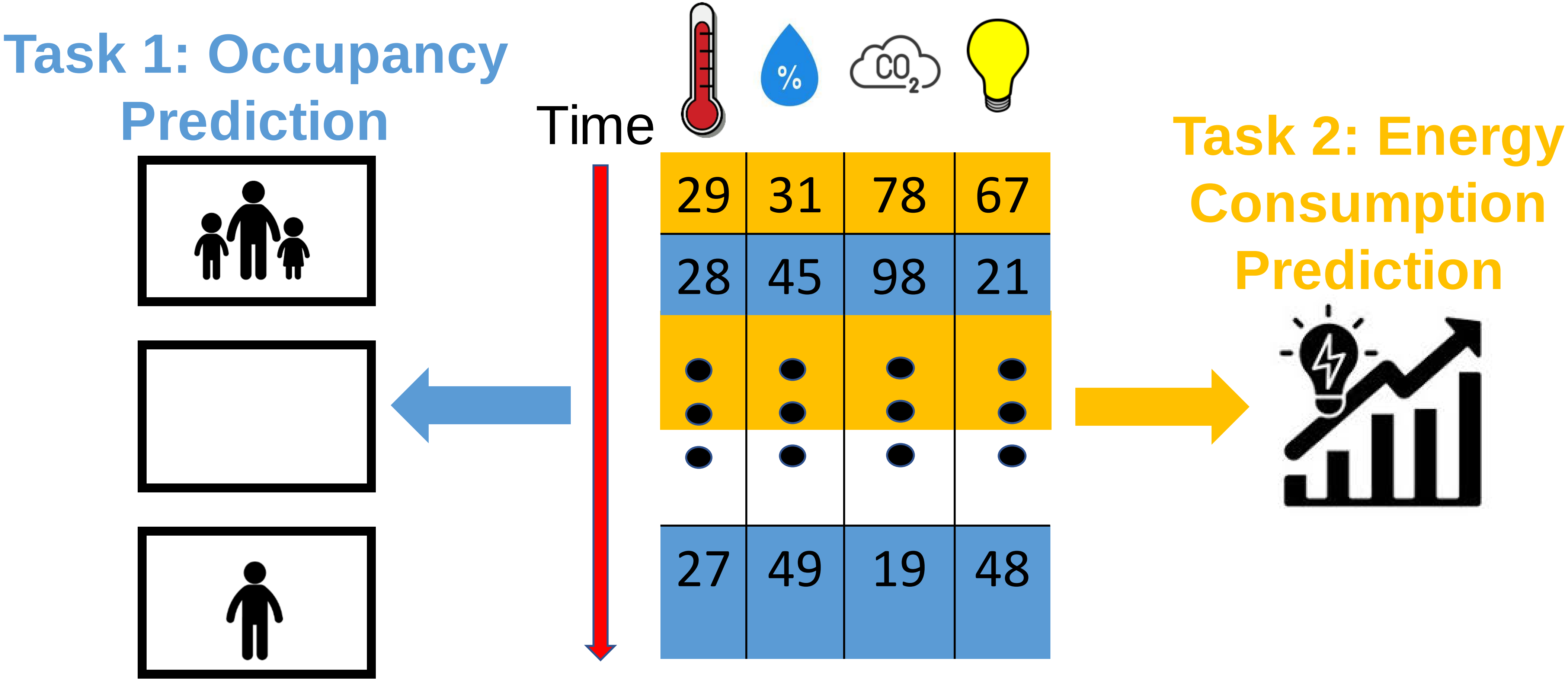
Task 1: Occupancy Prediction



Motivation


Time
29 31 78 67
28 45 98 21
27 49 19 48
Task 2: Energy Consumption Prediction

Motivation


Task 1: Occupancy Prediction
Time
CO_2
29 31 78 67
28 45 98 21
27 49 19 48
Task 2: Energy Consumption Prediction

Outline

- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

Related Work

- Statistical Methods: Distance-based[1], Shapelets[2], ROCKET[3]

[1] Abhilash Dorle, Fangyu Li, Wenzhan Song, and Sheng Li. 2020. Learning Discriminative Virtual Sequences for Time Series Classification. In CIKM. 2001–2004.
[2] Lexiang Ye and Eamonn Keogh. 2009. Time series shapelets: a new primitive for data mining. In KDD. 947–956.
[3] Angus Dempster, François Petitjean, and Geoffrey I Webb. 2020. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Mining and Knowledge Discovery 34, 5 (2020), 1454–1495.

Related Work

- Statistical Methods: Distance-based, Shapelets, ROCKET
- Deep Learning Methods:
  - Using labeled data: CNN[1], LSTM[2], Attention[3]

[1] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. 2019. Multivariate LSTM-FCNs for time series classification. Neural Networks 116 (2019), 237–245.
[2] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J Leon Zhao. 2014. Time series classification using multi-channels deep convolutional neural networks. In International conference on web-age information management. Springer, 298–310.
[3] Xuchao Zhang, Yifeng Gao, Jessica Lin, and Chang-Tien Lu. 2020. Tapnet: Multivariate time series classification with attentional prototypical network. In AAAI.

Related Work

- Statistical Methods: Distance-based, Shapelets, ROCKET
- Deep Learning Methods:
  - Using labeled data: CNN, LSTM, Attention
  - Using both unlabeled and labeled data: Negative Sampling[1], Contrastive Loss[2], Data Reconstruction[3]

[1] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. 2019. Unsupervised scalable representation learning for multivariate time series. arXiv preprint arXiv:1901.10738 (2019).
[2] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, and Bixiong Xu. 2021. Learning Timestamp-Level Representations for Time Series with Hierarchical Contrastive Loss. arXiv preprint arXiv:2106.10466 (2021).
[3] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. [n. d.]. A Transformer-based Framework for Multivariate Time Series Representation Learning. In KDD, pages=2114–2124, year=2021.

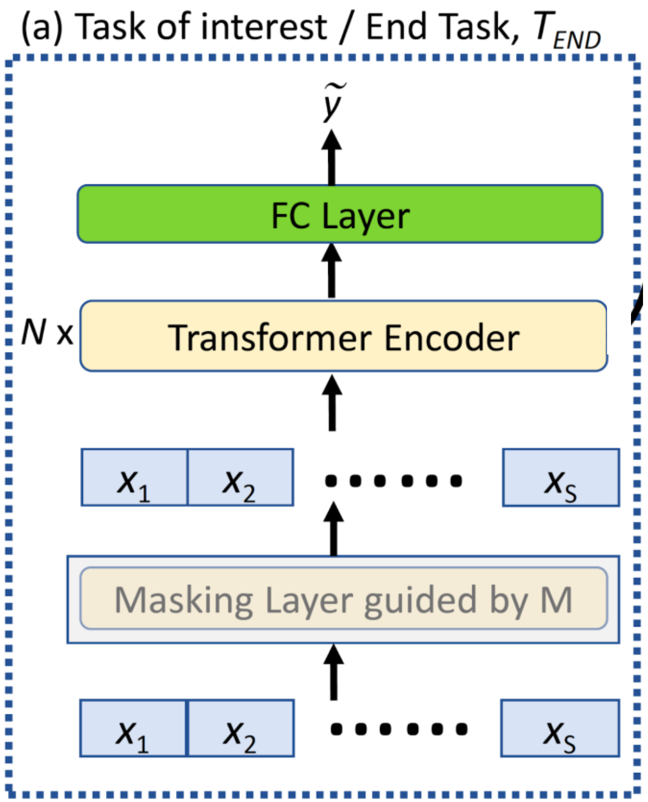
Outline

- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

Proposed Method

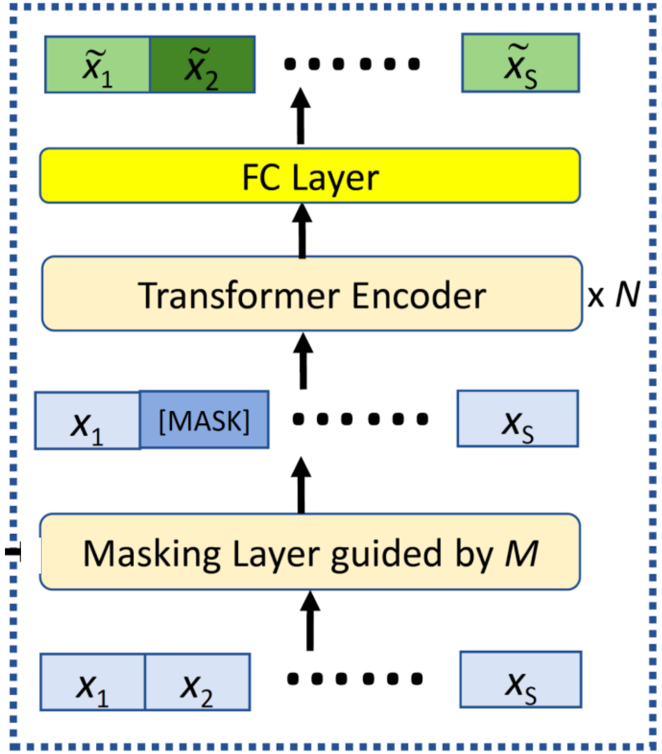
- Input: Uni-/multi-variate time-series X, Output: label y
- We use Transformer Encoder as the backbone model
- 3 modules
  a) End-task, T_{END}
  b) Data-driven Masking Strategy, M
  c) Task-aware Data Reconstruction, T_{TAR}

Proposed Method - T_{END}

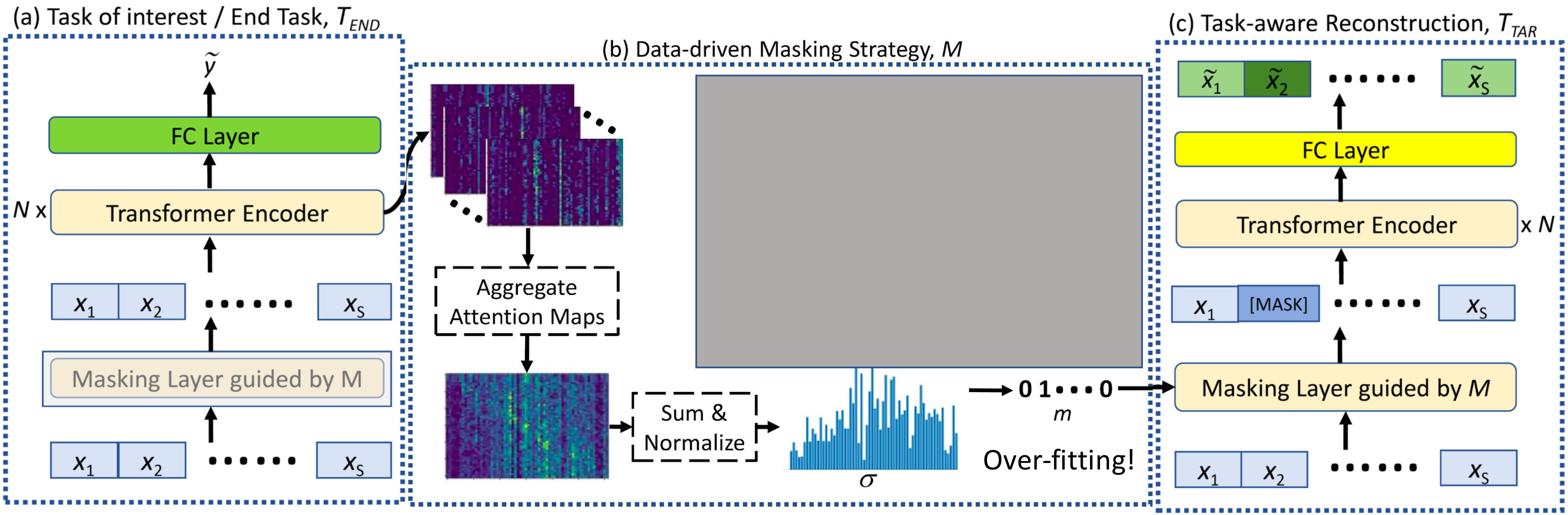


Proposed Method - T_{TAR}

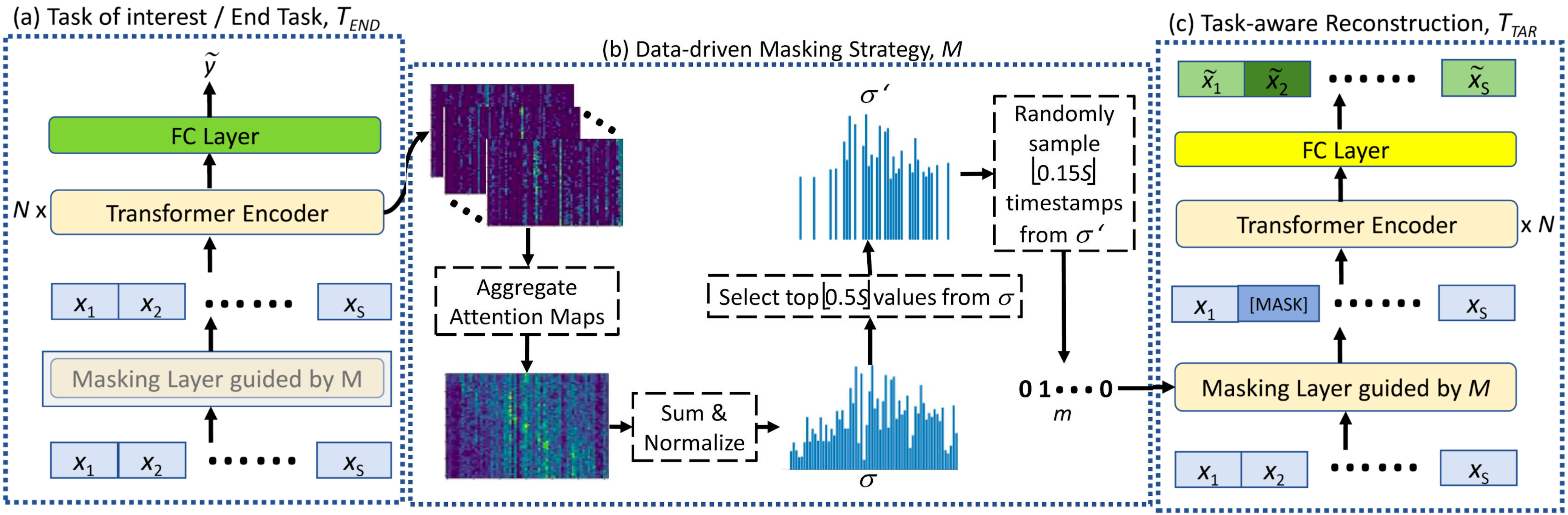
(c) Task-aware Reconstruction, T_{TAR}



Proposed Method



Proposed Method



Proposed Method

- T_{END} generates attention scores that is fed to M
- M selects a set of most important timestamps, and randomly samples a subset of those times to produce m
- Generated mask m decides which timestamps to mask during reconstruction, T_{TAR}

$$\mathcal{L}_{Total} = \eta \mathcal{L}_{TAR} + (1 - \eta) \mathcal{L}_{END}$$

Proposed Method

- Transformer Encoder shared by the two tasks
- The fully connected layers are task-specific
- T_{END} and T_{TAR} are trained in a multitask fashion

Outline

- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

Experimental Results - Classification

- Evaluated 34 classification datasets[1] across 14 baselines
- 2.7% higher average accuracy,
- 1.74-point lower average rank, and
- best results on 17 datasets compared to 7,
- by 2^{nd} best baseline TST[2]

[1] Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. 2018. The UEA multivariate time series classification archive, 2018. arXiv preprint arXiv:1811.00075 (2018).
[2] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. [n. d.]. A Transformer-based Framework for Multivariate Time Series Representation Learning. In KDD, pages=2114–2124, year=2021.

Experimental Results - Regression

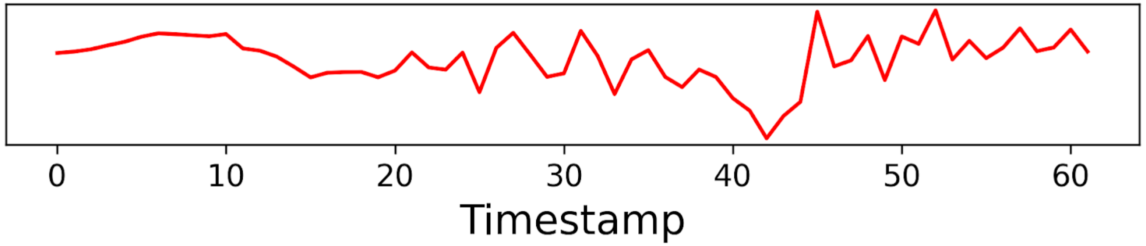
- Evaluated 6 regression datasets[1] across 12 baselines
- 0.67-point lower average rank, and
- best results on 3 datasets compared to 2,
- by 2nd best baseline TST[2]

[1] Chang Wei Tan, Christoph Bergmeir, Francois Petitjean, and Geoffrey I Webb. 2020. Monash university, uea, ucr time series regression archive. arXiv e-prints (2020), arXiv–2006.
[2] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. [n. d.]. A Transformer-based Framework for Multivariate Time Series Representation Learning. In KDD, pages=2114–2124, year=2021.

Outline

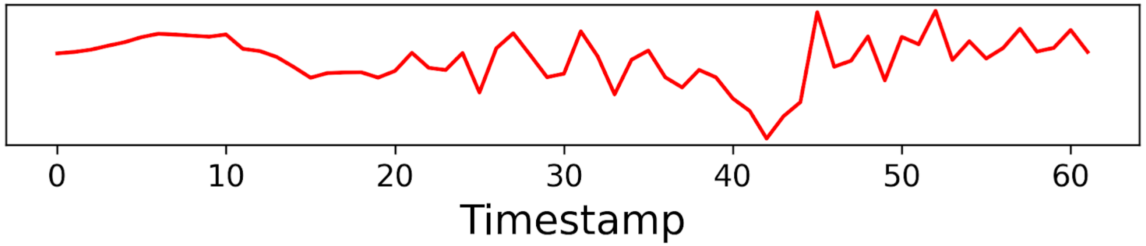
- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

Case Study – Face Detection

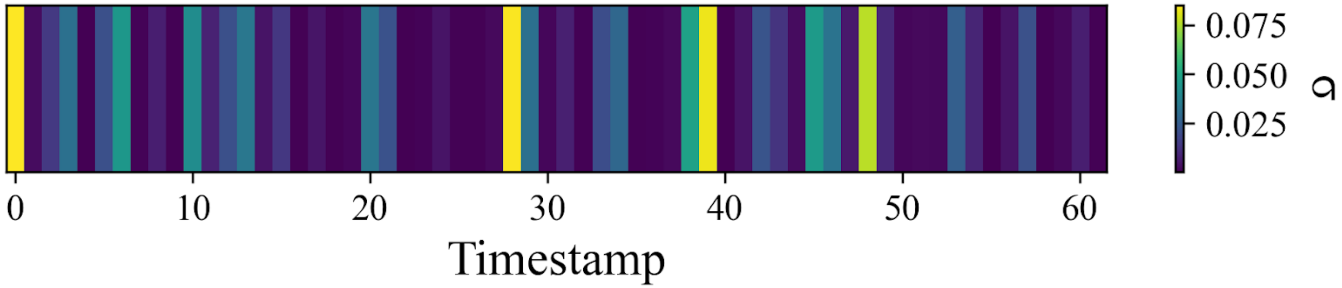


Time Series Data

Case Study – Face Detection

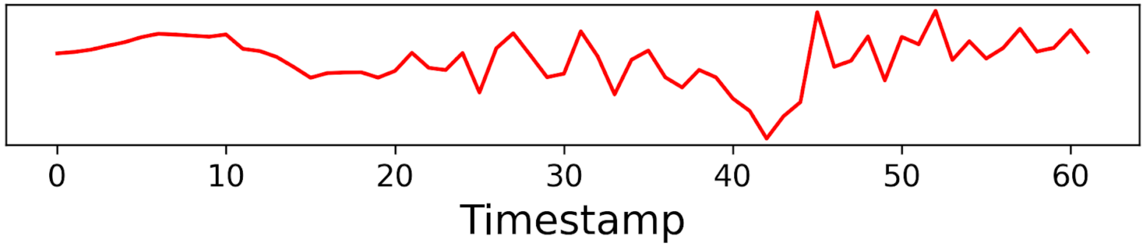


Time Series Data

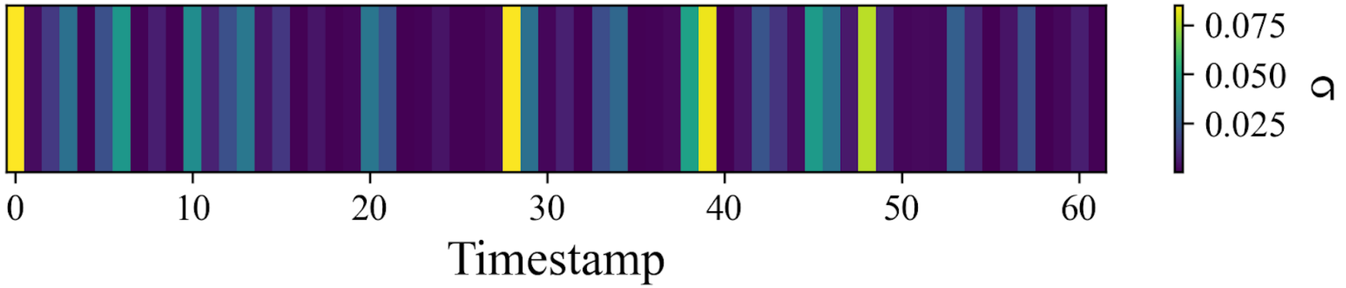


Random Data Reconstruction

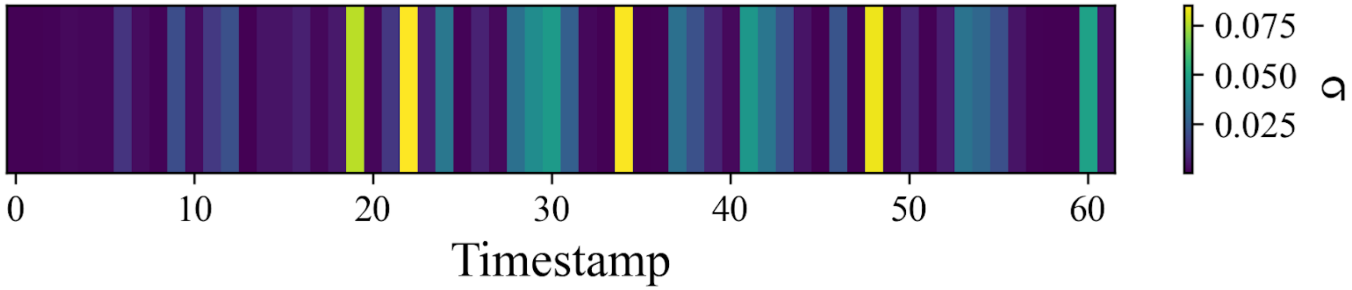
Case Study – Face Detection



Time Series Data



Random Data Reconstruction



Task-aware Data Reconstruction

Outline

- Motivation
- Related Work
- Proposed Method
- Experimental Results
- Case Study
- Conclusion

Conclusion

- Task-agnostic SSL may produce sub-optimal performance
- Learn *task-aware* data representation customized to the end-task
- End-task and reconstruction task trained alternately
- Data-driven masking strategy uses attention scores to find timestamps deemed important by end-task and mask them for reconstruction
- TARNet outperforms 26 baselines on 40 datasets
- Case study shows task-aware method captures domain-specific inherent properties from data

Questions?

- Contact: Ranak Roy Chowdhury (rrchowdh@eng.ucsd.edu)
- Implementation: https://github.com/ranakroychowdhury/TARNet
- Paper: https://dl.acm.org/doi/10.1145/3534678.3539329