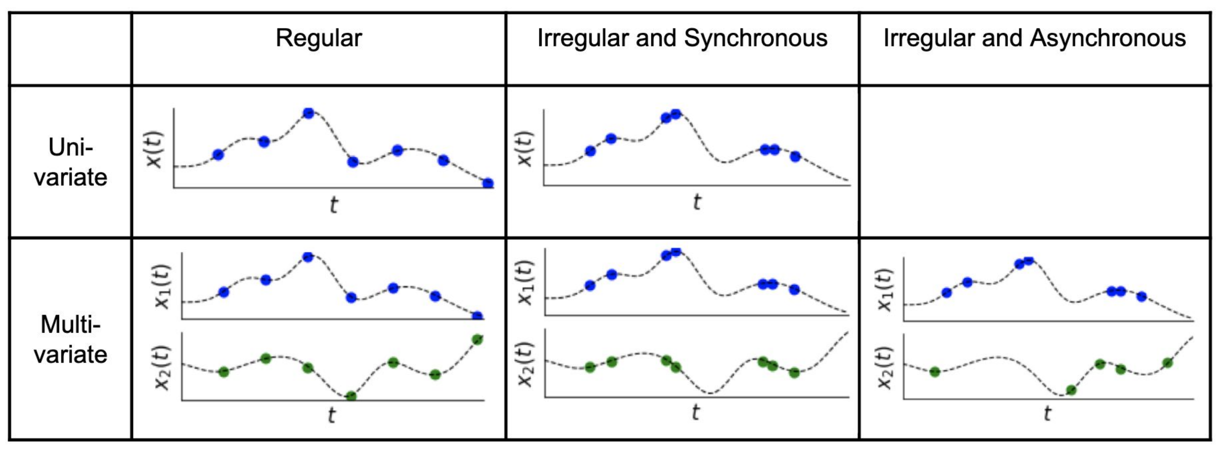
PrimeNet: Pre-Training for Irregular Multivariate Time Series

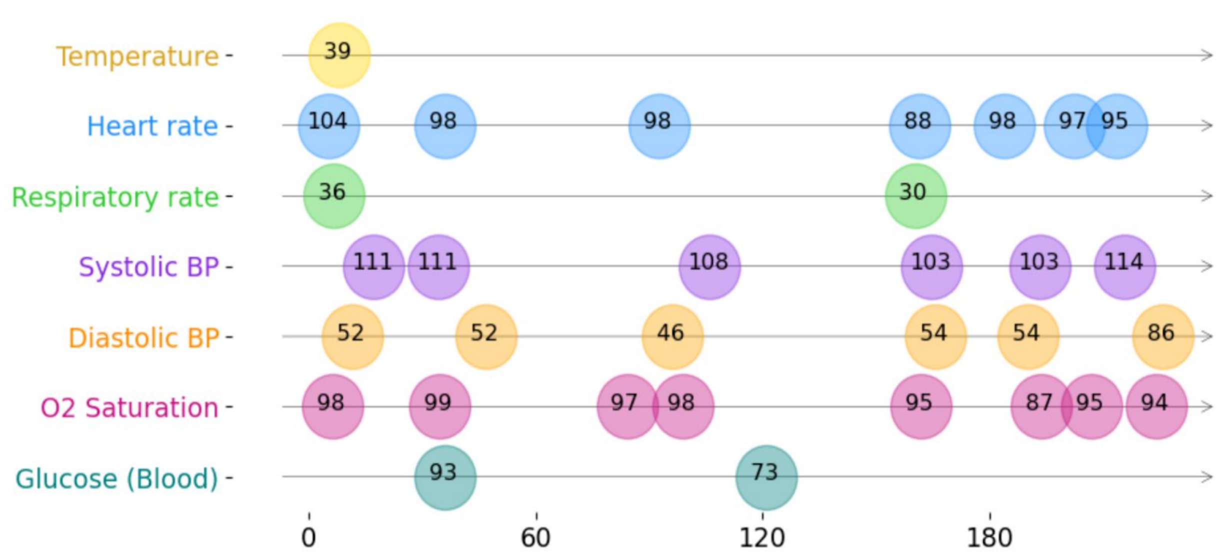
Ranak Roy Chowdhury,[1] Jiacheng Li,[1] Xiyuan Zhang,[1] Dezhi Hong,[2] Rajesh Gupta,[1] Jingbo Shang[1]

{rrchowdh, j9li, gupta, jshang}@eng.ucsd.edu, xiyuanzh@ucsd.edu, hondezhi@amazon.com

[1]University of California San Diego, [2]Amazon

Motivation

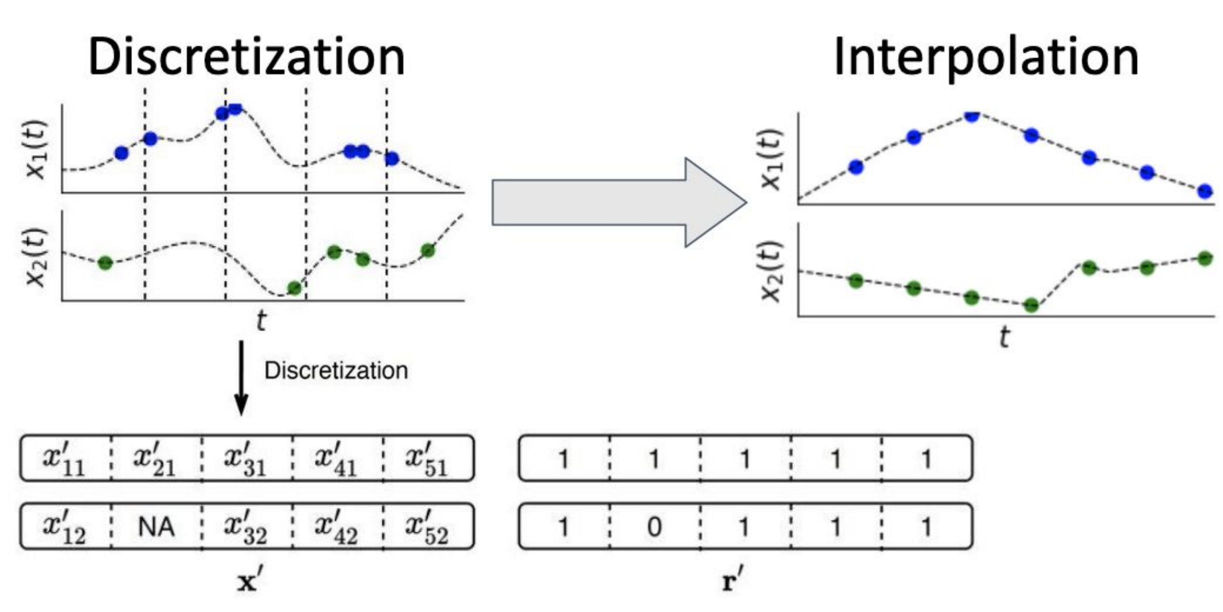




minutes from ICU admission

Current Approaches

- 2 approaches to learn from *irregular* data:
  - Discretization followed by interpolation
  - Explicitly model irregularity - Neural ODE, Set, Time Attention



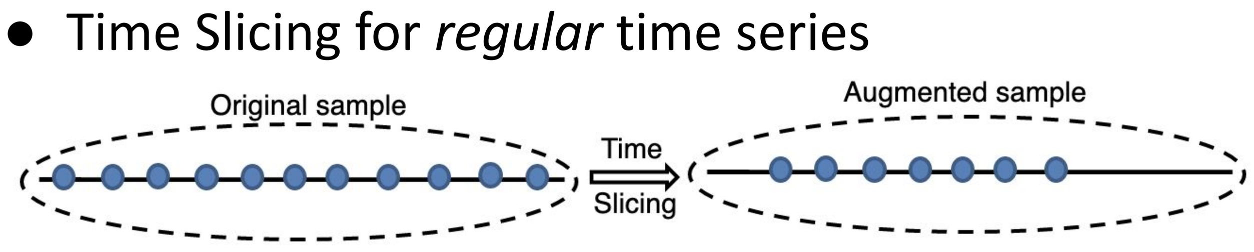
- Fixing the bin size is problematic:
  - Narrow bins: high missing %, sequence explodes
  - Wide bins - data aggregation, lose of fine-grained details
- Ignores irregularity pattern that may be useful to end task, harming performance
- Neural ODE, Set, Time Attention take sampling time as input and explicitly encode irregularity

Objective

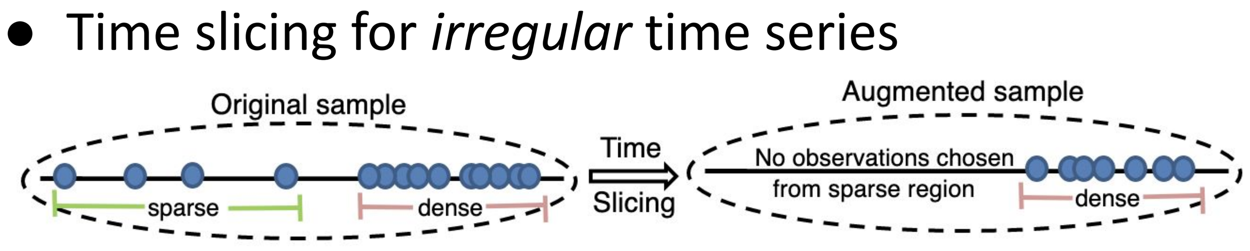
- Self-supervised models need regular intervals.
- Irregular models are fully- or semi-supervised.
- Goal: Learn fully self-supervised representation from unlabeled irregular multivariate time series?
- Two self-supervision tasks for irregular series:
  - Contrastive Learning
  - Masked Reconstruction

Time Contrastive Learning (TimeCL)

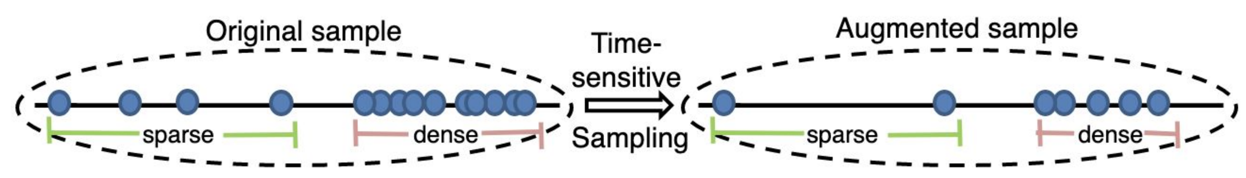
- Time Slicing for *regular* time series



- Time slicing for *irregular* time series



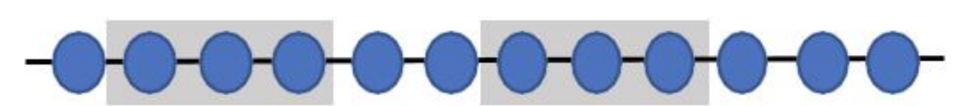
- Time-sensitive sampling for *irregular* time series



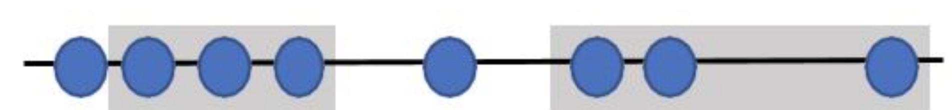
$$\mathcal{L}_{CL} = -\log \frac{\exp(\tilde{X}_i \tilde{X}_j/\tau)}{\sum_{k=1}^{2B} \exp(\tilde{X}_i \tilde{X}_k/\tau)}$$

Time Reconstruction (TimeReco)

- Constant length masking for regular series



- Constant length masking for irregular series



- Constant time masking for irregular series

$$\mathcal{L}_{Reco} = \left\| \mathrm{M}_V \odot (\tilde{X}_U - X_V) \right\|_2^2$$

$$\mathcal{L} = \eta \mathcal{L}_{CL} + (1-\eta)\mathcal{L}_{Reco}$$

PrimeNet

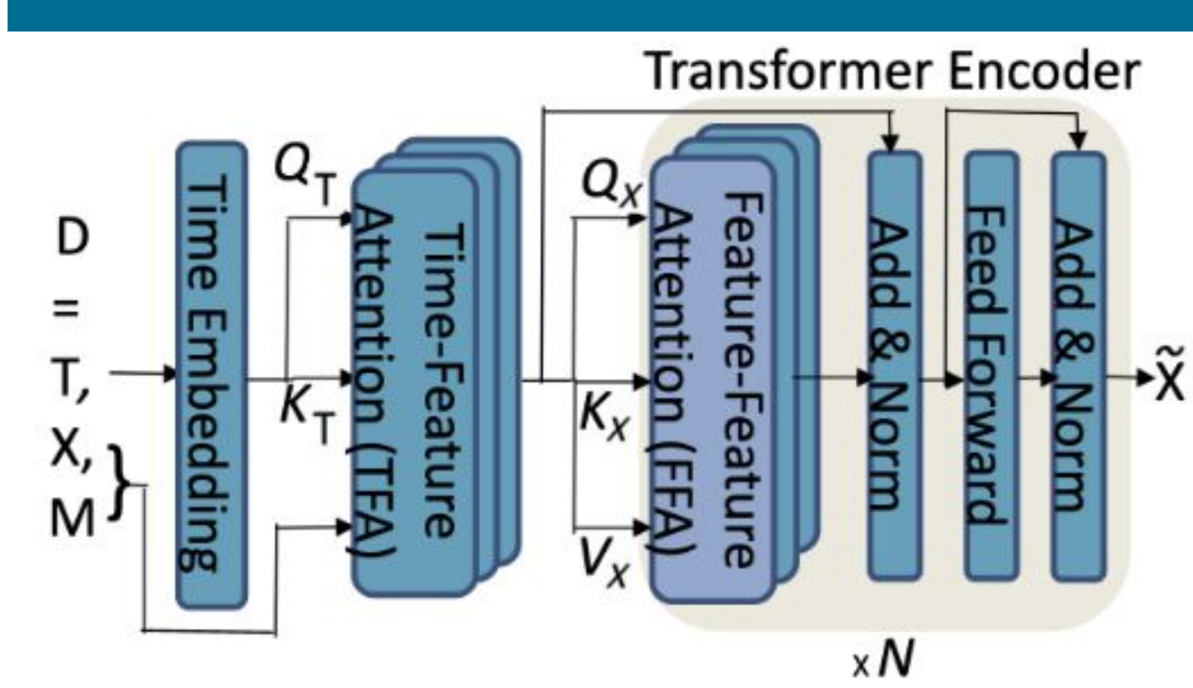


Fig: Model Architecture

- Time-Embedding Layer (Shukla 21)

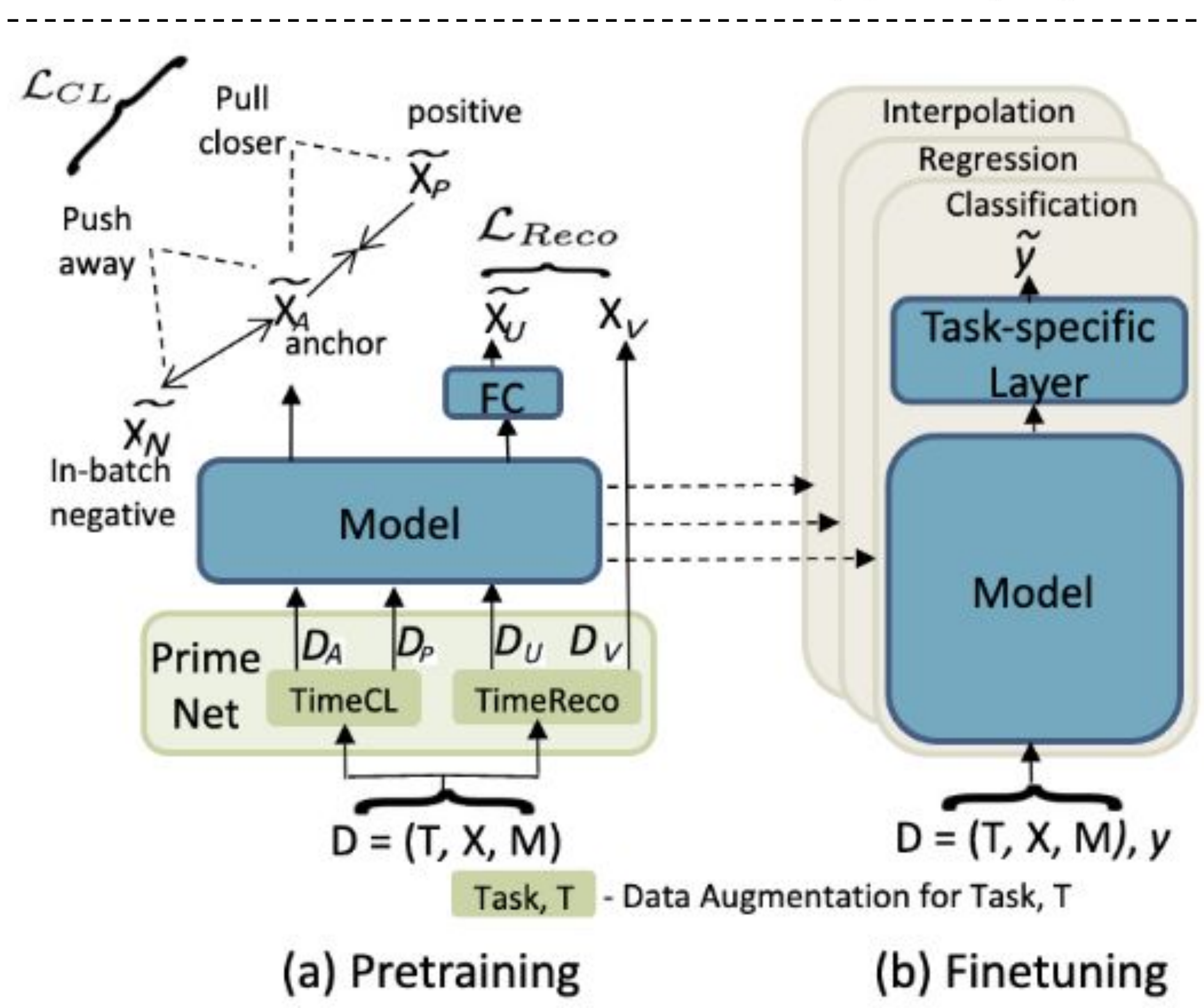
$$\phi h(\mathrm{T})[i] = \begin{cases} \omega_{0h}.\mathrm{T} + \alpha_{0h}, \text{ if } i = 0 \\ sin(\omega_{ih}.\mathrm{T} + \alpha_{ih}), \text{ if } 0 < i < d_r \end{cases}$$

- Time-Feature Attention (TFA) (Shukla 21)

$$\mathrm{TFA}(\mathrm{Q}_T, \mathrm{K}_T, \mathrm{M}, \mathrm{X}) = (\mathrm{M} \odot \mathrm{A}_T)\mathrm{X},$$
$$\mathrm{A}_T = \mathrm{softmax}(\mathrm{Q}_T \mathrm{K}_T / d_r)$$

- Feature-Feature Attention (FFA) (Vaswani 17)

$$\mathrm{FFA}(\mathrm{Q}_X, \mathrm{K}_X, \mathrm{V}_X, \mathrm{M}) = (\mathrm{M} \odot \mathrm{A})\mathrm{V}_X,$$
$$\mathrm{A} = \mathrm{softmax}(\mathrm{Q}_X \mathrm{K}_X / d_r)$$



(a) Pretraining (b) Finetuning

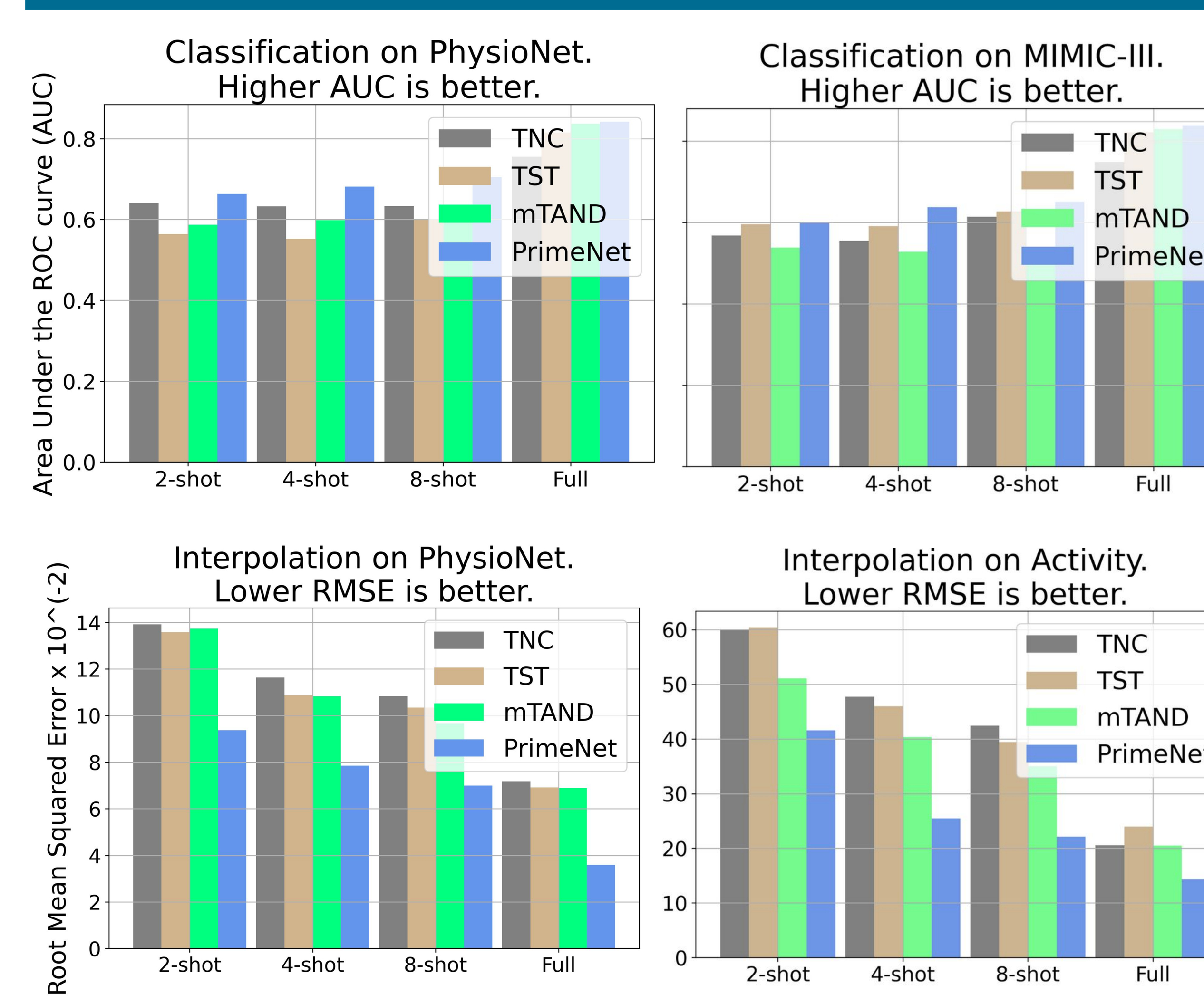
Code is publicly available at:
https://github.com/ranakroychowdhury/PrimeNet

Datasets and Baselines

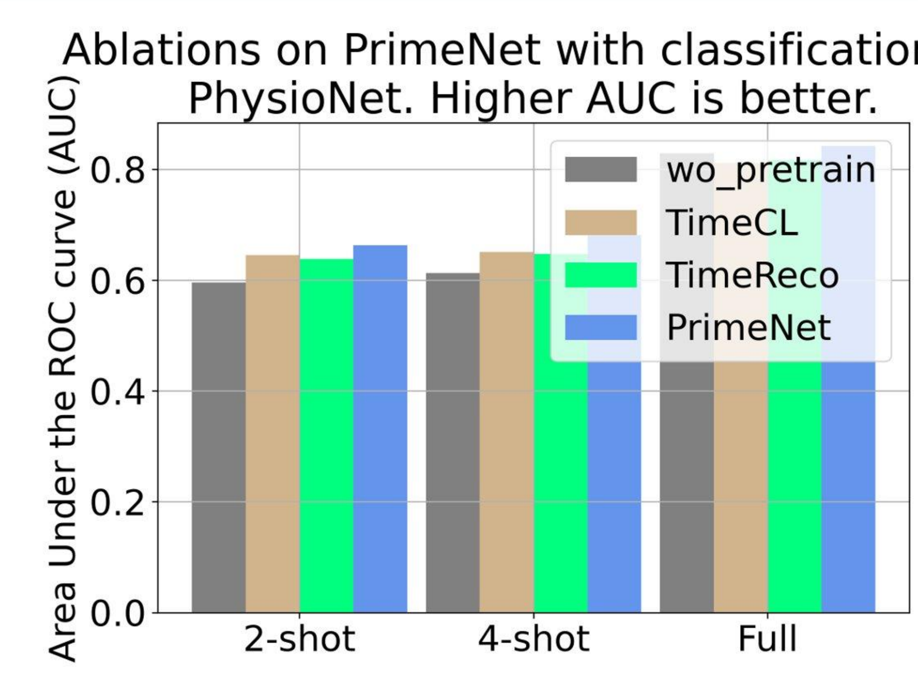
- Real-world irregular data - Healthcare, IoT
- 48 hours of physiological data of ICU patients in *PhysioNet* and *MIMIC-III* datasets
- 3D position of human body from *Activity* dataset
- Weather data from *Appliances Energy* dataset
- Missing ratio statistics (mean, std dev): Physio Net (0.86, 0.24), MIMIC-III (0.65, 0.36), Activity (0.75, 0.64), Appliances Energy (0.87, 0.47)

- Self-supervised regular model: TS2Vec (Yue 21), TNC (Tonekaboni 21), TST (Zerveas '21)
- Fully- or semi-supervised irregular model: GRU-Mean (Che 18), P-LSTM (Neil 16), RNN-VAE(Chen 18), ODE-RNN (Rubanova 19), L-ODE (Rubanova 19), mTAND (Shukla 21)

Results



Ablation



Conclusion

- Pre-trained model for irregular time series
- 2 time-sensitive self-supervision tasks:
  - TimeCL: preserve original sampling density while augmenting irregular samples
  - TimeReco: adjust mask length based on sampling density to solve reconstruction
- Pre-trained model fine-tuned on labeled data
- Beats SOTA on Healthcare, IoT data for classification, interpolation, regression for full and few-shot data setting
- Ablation shows essence of both tasks and design choices