VIRGINIA TECH™

Synthetic Time-Series Data Generation for Cyber-Physical Systems

Minh B. T. Nguyen
Virginia Tech
Blacksburg, VA 24060
mnguyen0226@vt.edu

Md Nazmul Kabir Sikder
Virginia Tech
Arlington, VA 22203
nazmulkabir@vt.edu

Chelsea Wang
Virginia Tech
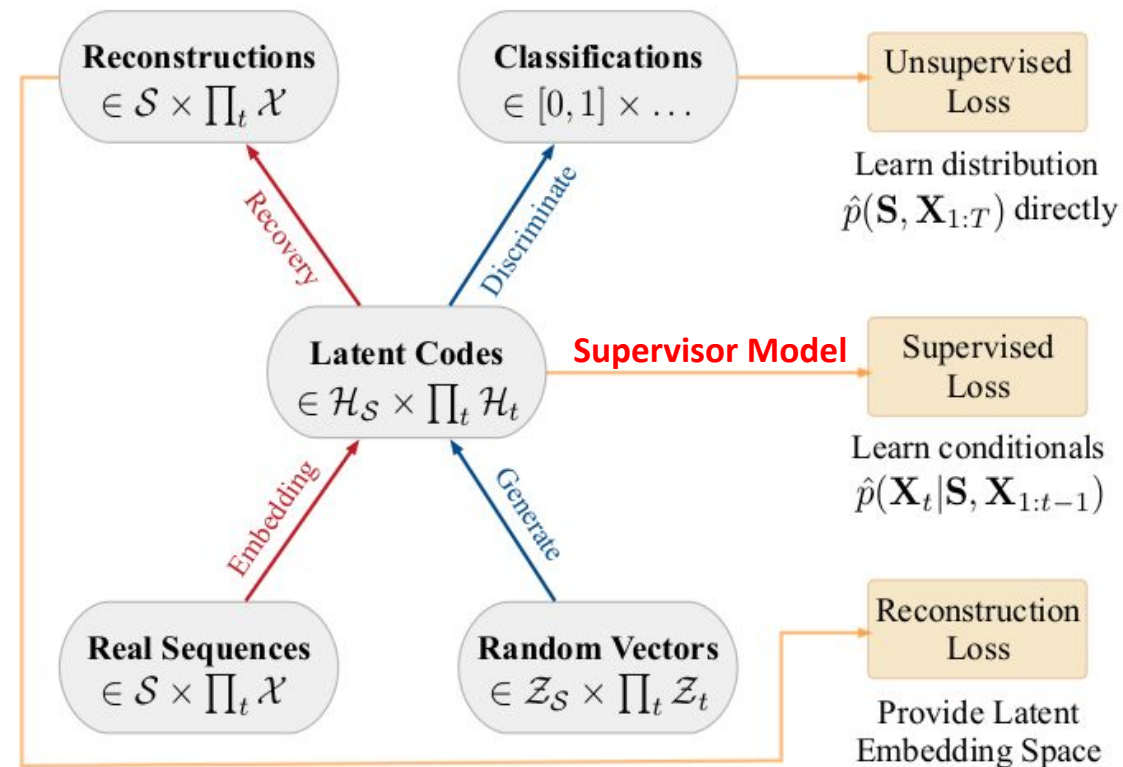Arlington, VA 22203
chelseawang@vt.edu

1. TimeGAN

4 Components:

- Autoencoder (Embedding, Recovery), Generator, Discriminator, (and Supervisor).

3 Training Phases:

- Autoencoder unsupervised training.

- Supervisor supervised training.

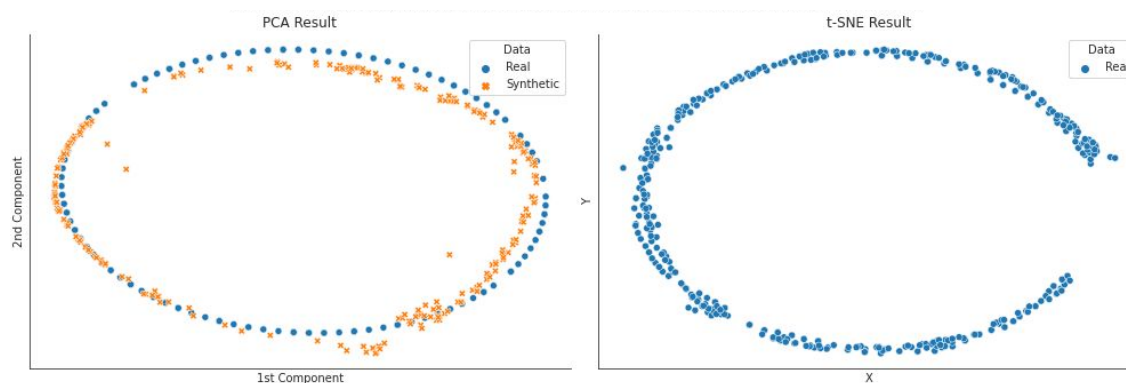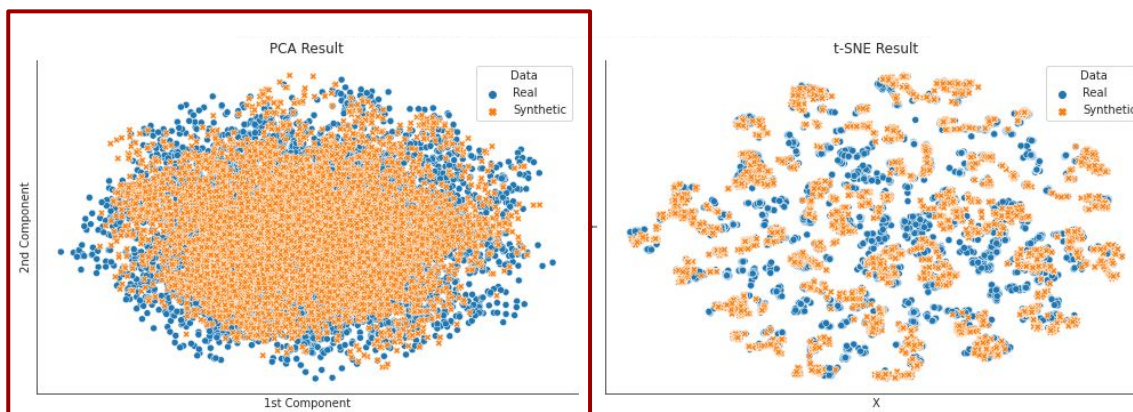- Joint Training with Encoder, Generate and Iterate.

Reconstructions $\in \mathcal{S} \times \prod_t \mathcal{X}$

Classifications $\in [0,1] \times \ldots$

Unsupervised Loss

Learn distribution $\hat{p}(\mathbf{S}, \mathbf{X}_{1:T})$ directly

Recovery

Discriminate

Latent Codes $\in \mathcal{H}_{\mathcal{S}} \times \prod_t \mathcal{H}_t$

Supervisor Model

Supervised Loss

Learn conditionals $\hat{p}(\mathbf{X}_t | \mathbf{S}, \mathbf{X}_{1:t-1})$

Embedding

Generate

Real Sequences $\in \mathcal{S} \times \prod_t \mathcal{X}$

Random Vectors $\in \mathcal{Z}_{\mathcal{S}} \times \prod_t \mathcal{Z}_t$

Reconstruction Loss

Provide Latent Embedding Space

Yoon, Jinsung, Daniel Jarrett, and Mihaela Van der Schaar. "Time-series generative adversarial networks." *Advances in Neural Information Processing Systems* 32 (2019).

2. Experiment

❏ Model Training and Data Generation
 ❑ Sine Wave.
 ❑ BATADAL (WDS dataset).
 ❑ SWaT (WTS dataset).
❏ Evaluation
 ❑ Visualization: PCA, t-SNE.
 ❑ Discriminative Score: Real vs Fake Classifier.
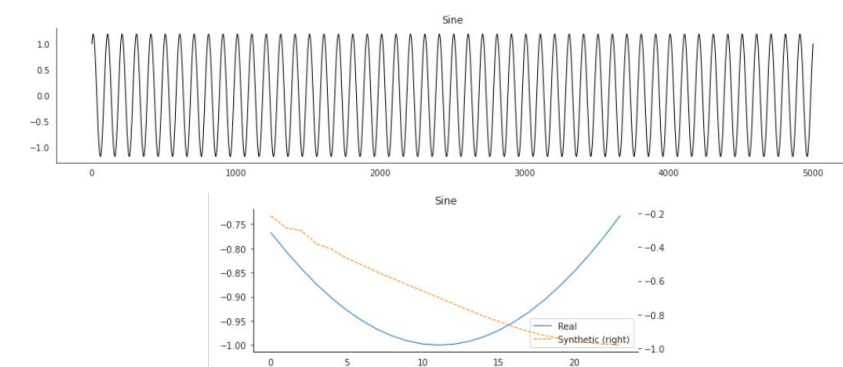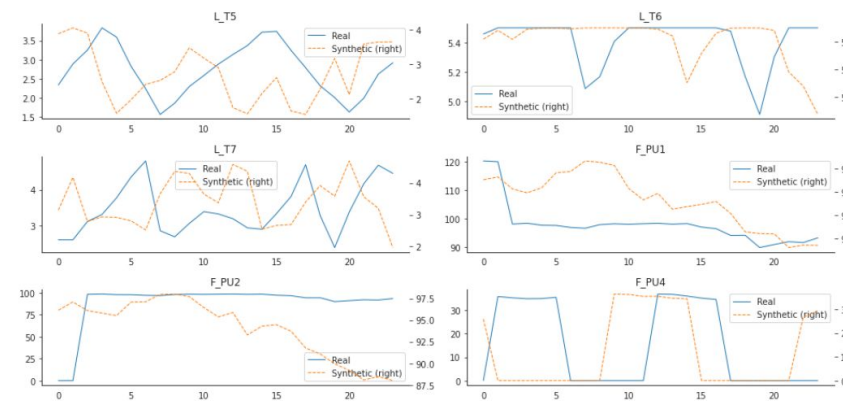 ❑ Predictive Score: TRTR, TRTS, TSTR, TSTS.

Yoon, Jinsung, Daniel Jarrett, and Mihaela Van der Schaar. "Time-series generative adversarial networks." *Advances in Neural Information Processing Systems* 32 (2019).
Goh, Jonathan, et al. "A dataset to support research in the design of secure water treatment systems." *International conference on critical information infrastructures security*. Springer, Cham, 2016.
Taormina, Riccardo, et al. "Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks." *Journal of Water Resources Planning and Management* 144.8 (2018): 04018048.
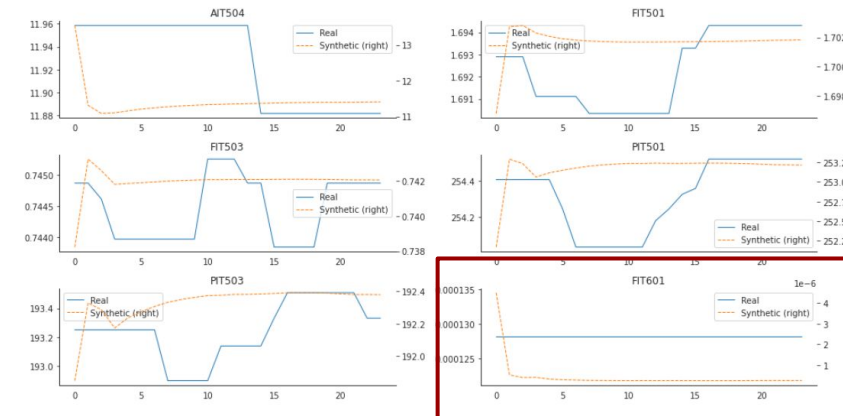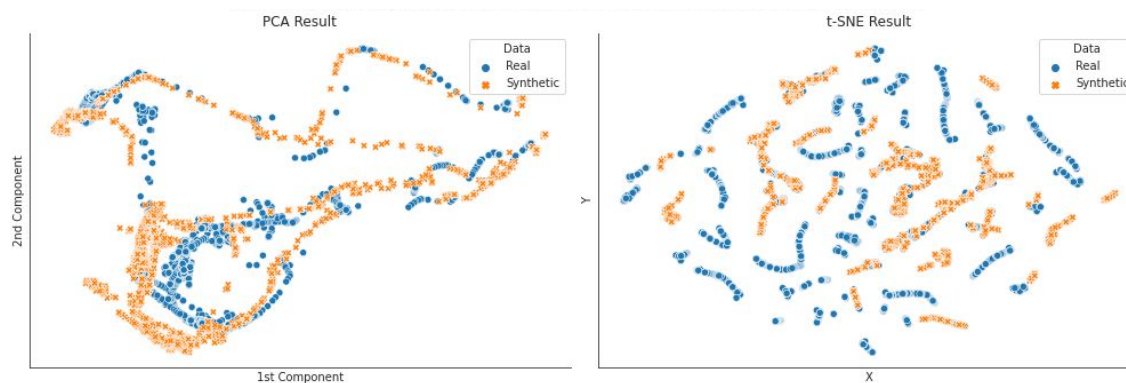
3. Results: Visualization

Sine Wave

BATADAL Dataset

SWaT Dataset
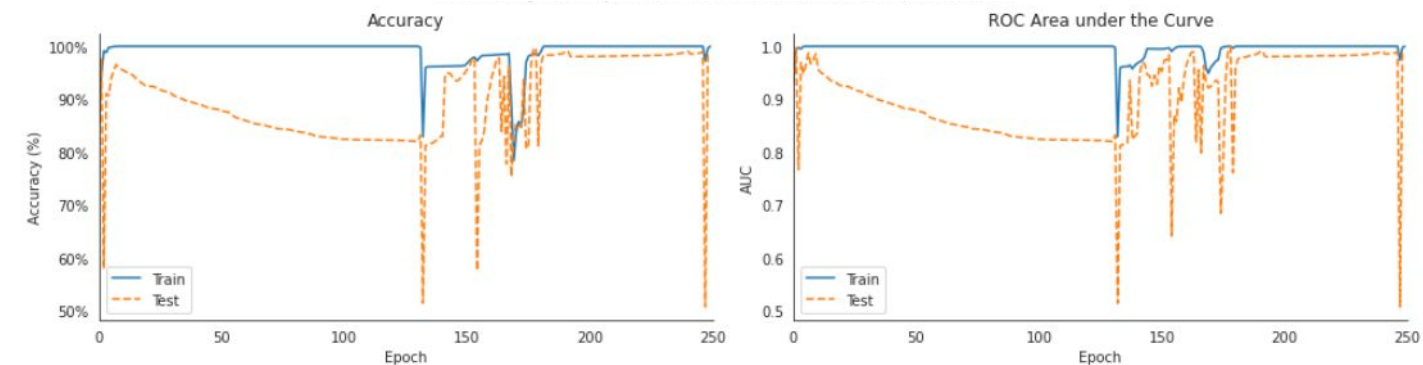
3. Results: Discriminative

Sine Wave

BATADAL Dataset
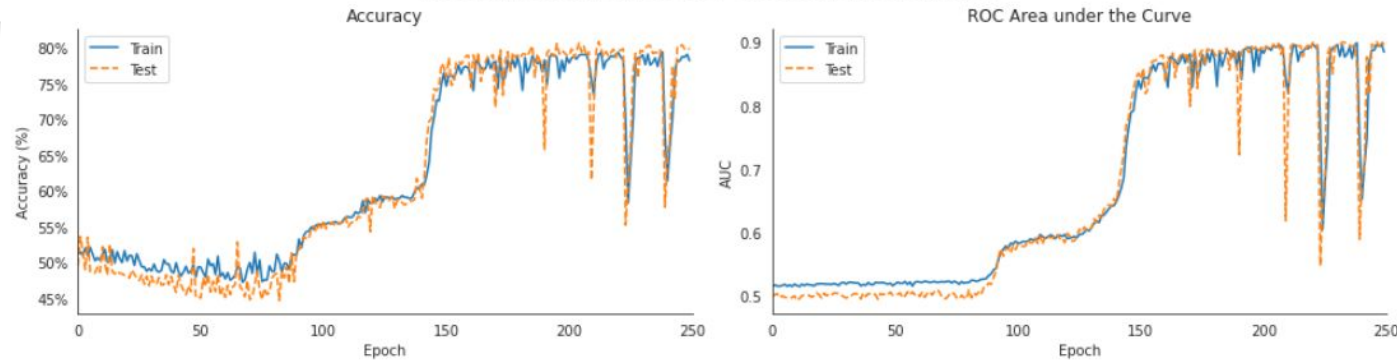
SWaT Dataset

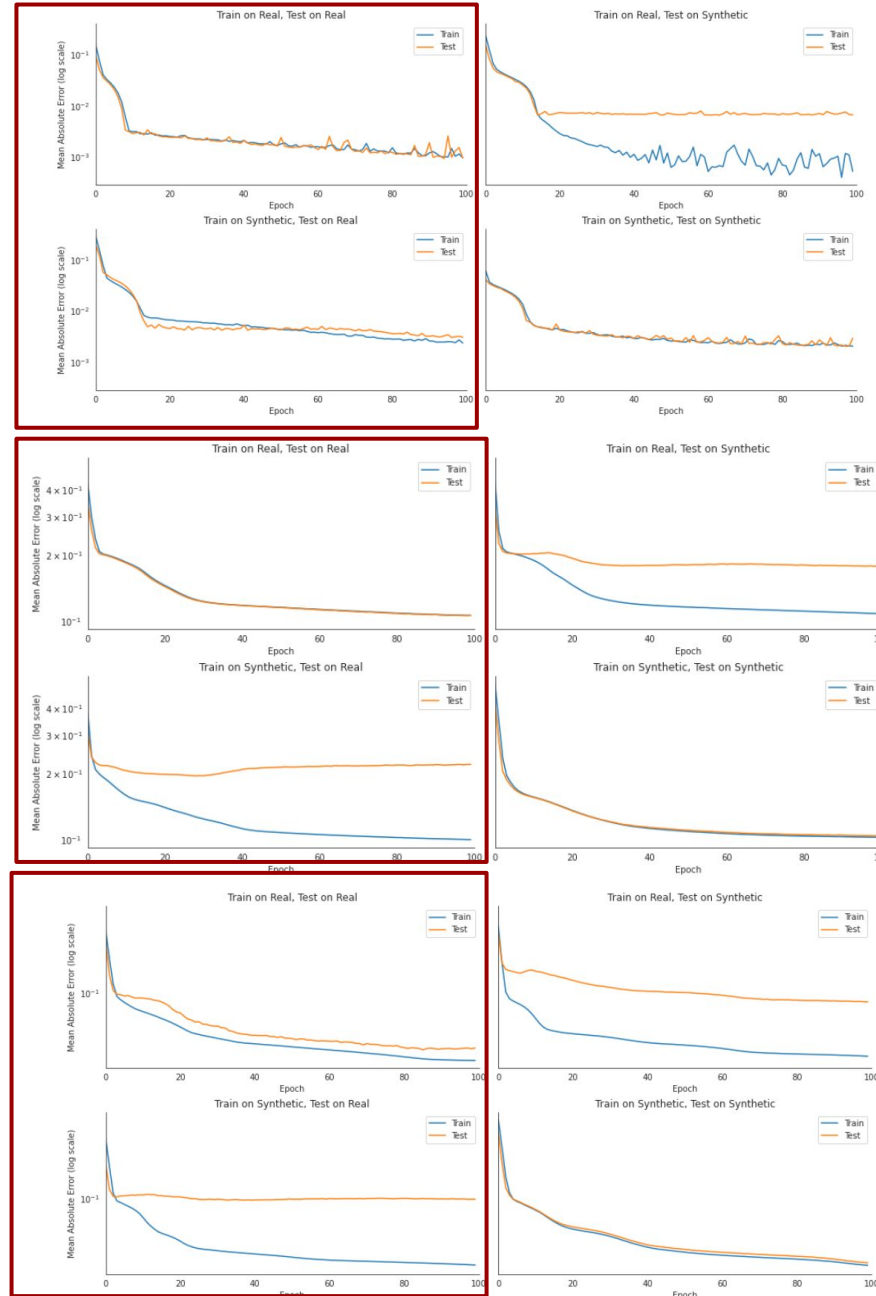|	AUC	Test Acc
Sine	0.8960	0.7980
BATADAL	0.98	0.98
SWaT	0.9876	0.9630

3. Results: Discriminative

Sine Wave

BATADAL Dataset

SWaT Dataset

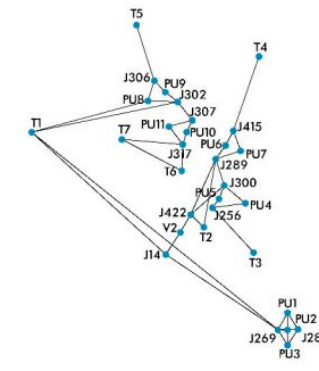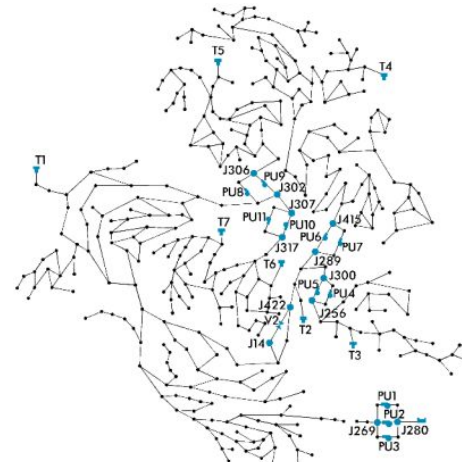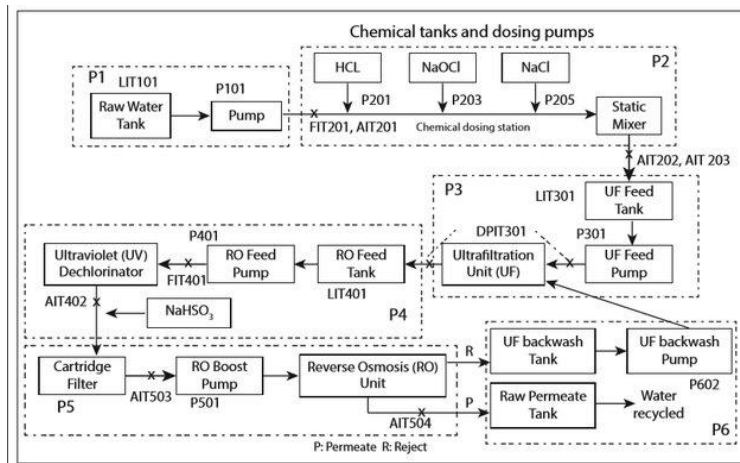| Four Tests |
|---|
| Train-Real, Test Real |
| Train-Real, Test Synthetic |
| Train-Synthetic, Test Synthetic |
| Train-Synthetic, Test Real |

4. Conclusion & Future Works

- Conclusions:
    - Time GANs was able to preserve temporal-dependency for multivariate time-series.
    - Synthetic dataset can replace the real dataset in both training and testing cases.
- Future Works:
    - Differential Privacy.
    - Label/Class-dependence dataset.
    - Spatio-Dependency Preservation - GNN/Dynamic Graph?

Goh, Jonathan, et al. "A dataset to support research in the design of secure water treatment systems." *International conference on critical information infrastructures security*. Springer, Cham, 2016.
Taormina, Riccardo, et al. "Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks." *Journal of Water Resources Planning and Management* 144.8 (2018): 04018048.

1. Background

Motivation:

- Open-source CPS datasets are limited.

- GANs is successful in research and domains.

Solution: GANs for Time-series data.

Hypothesis: "Can we generate quality synthetic time-series data?"

Literature Review:

Models for Time-Series Data	Capture Temporal Dependence	Continuous Variable
WP-GAN	Yes	No
CT-GAN	Partial	Yes
RCGAN	Partial	Yes
TimeGAN	Yes	Yes