Machine-learning based Synthetic Data Generation

Ning Lu

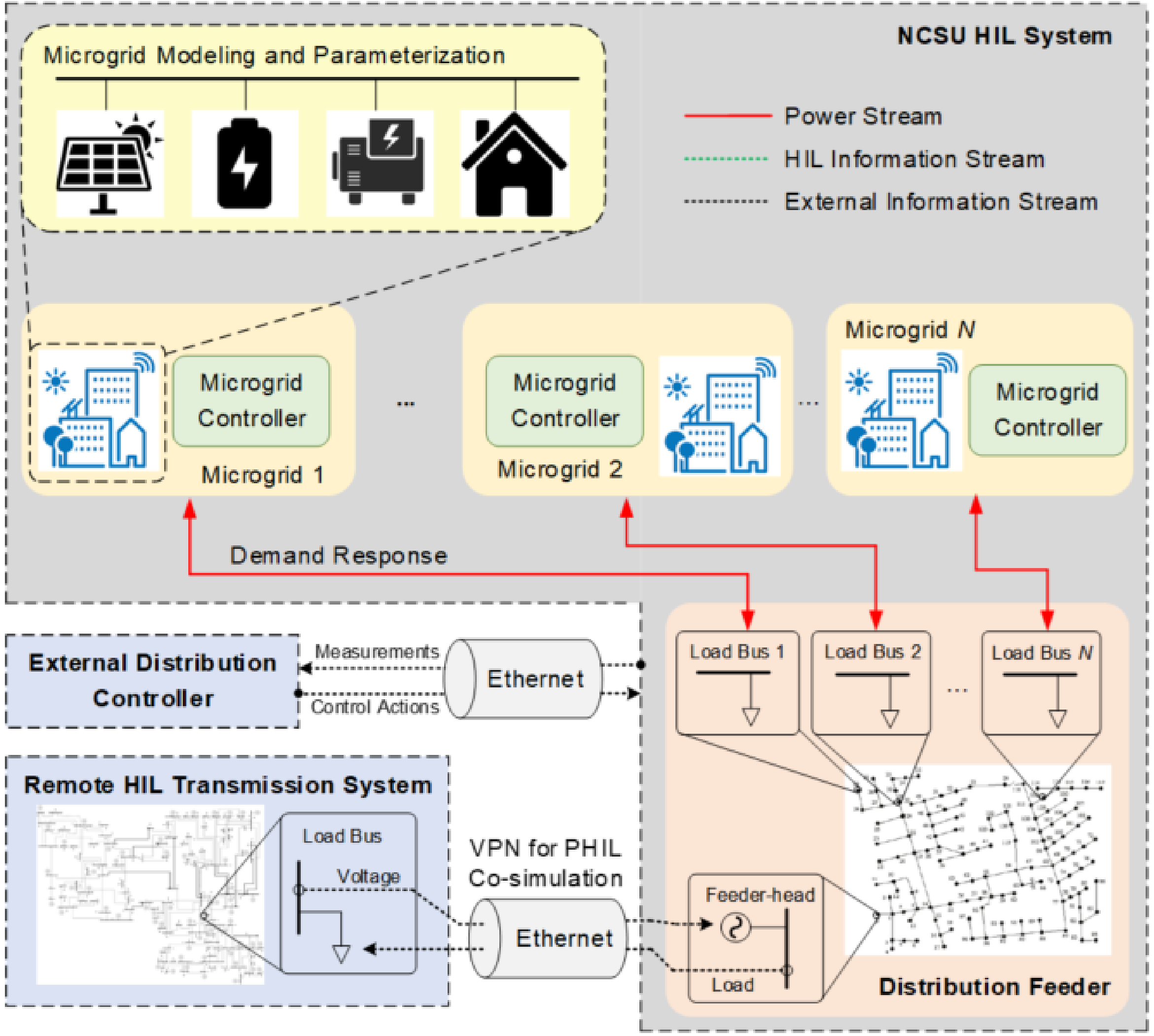
NC STATE UNIVERSITY

GridWrx Lab

Prepared for 2023 IEEE PES General Meeting

Outline

- Why do we need synthetic data?
  - Enable digital-twin based simulation
- Approach
  - Directly use real data
  - Statistic-based load profile generation
  - Ours: **Generative machine-learning based**
- Considerations
  - Realisticness
  - Customizable data resolution
  - Preserve temporal, spatial, group correlations
- Conclusions

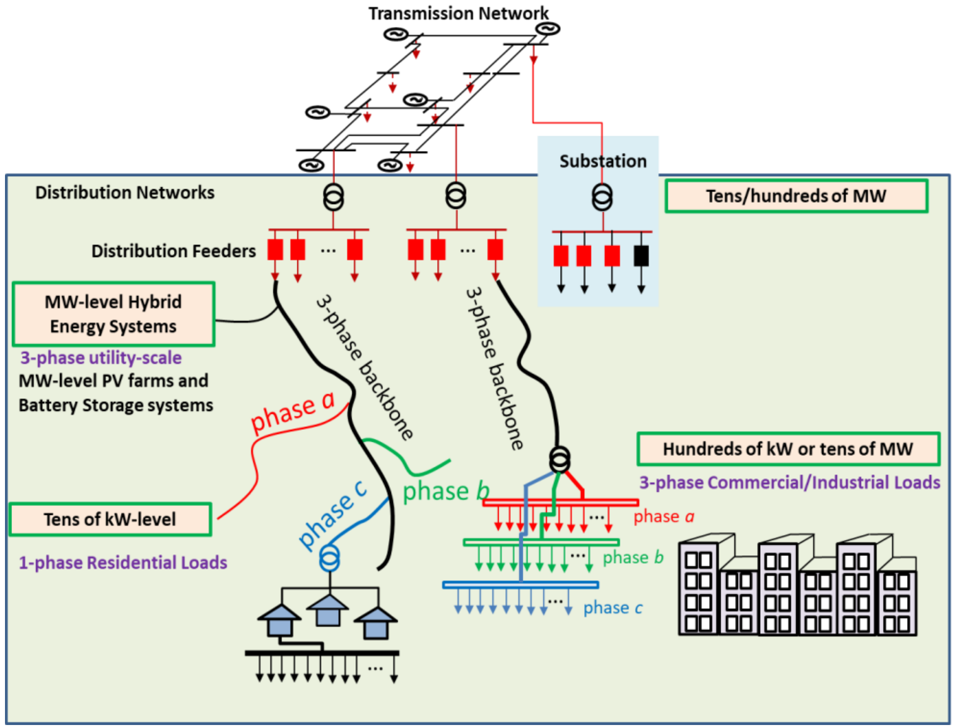


https://sites.google.com/a/ncsu.edu/ninglu/pars-platform?authuser=0

1. Using Real-data

Transforming low resolution data to high resolution

Data Resolution



1-Minute Sub-metered data
- End use consumptions of appliances
- Not usually available
- Enabling technologies: IoT sensors

Super resolution

15-minute Smart Meter Data
- Average kWH, kVar, Voltage
- Sensitive information

Hourly
- Temperature, irradiance
- Average kWH

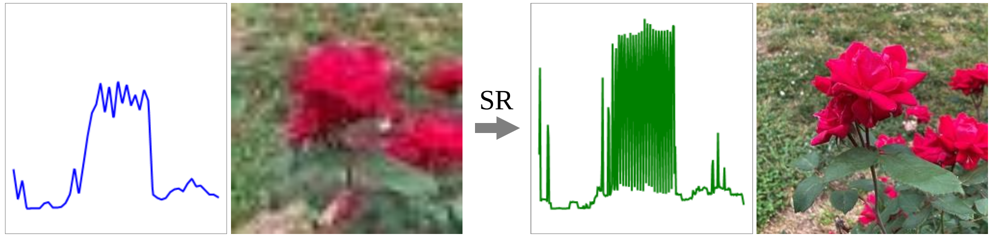
Daily
- Peak hour
- DR events

Monthly
- Utility billing information
- Peak day peak hour

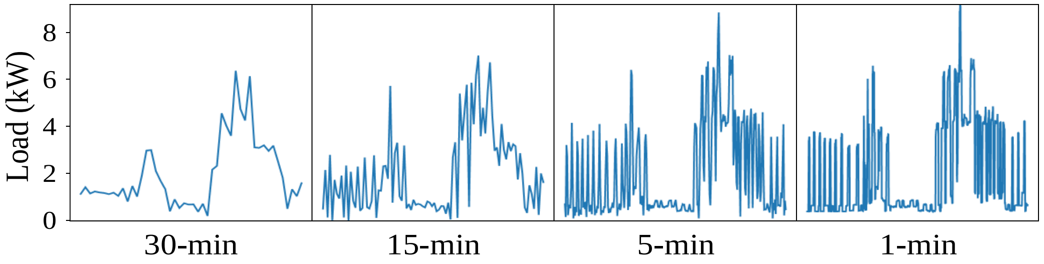
Time

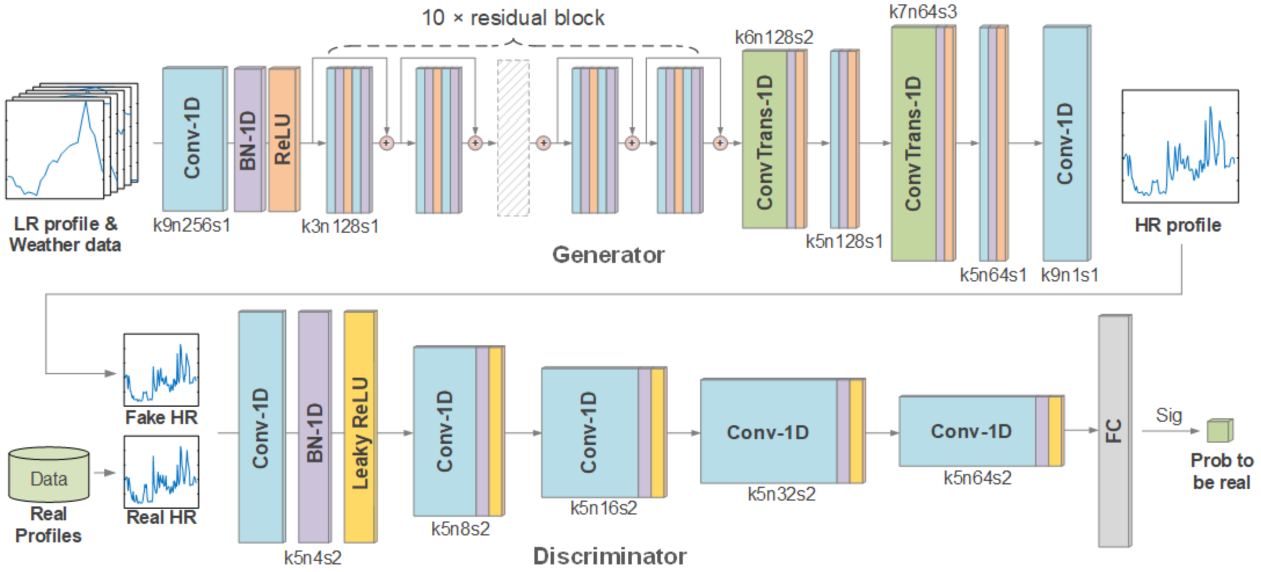
ProfileSR-GAN A GAN-based Super-resolution Method

- Develop high-resolution PV and load profiles
- Inputs: **15-min** or **30-min** low resolution (LR)
- Restore the high-frequency load dynamics from the LR measurements using deep learning methods



LR profile Blurred image HR profile HR image





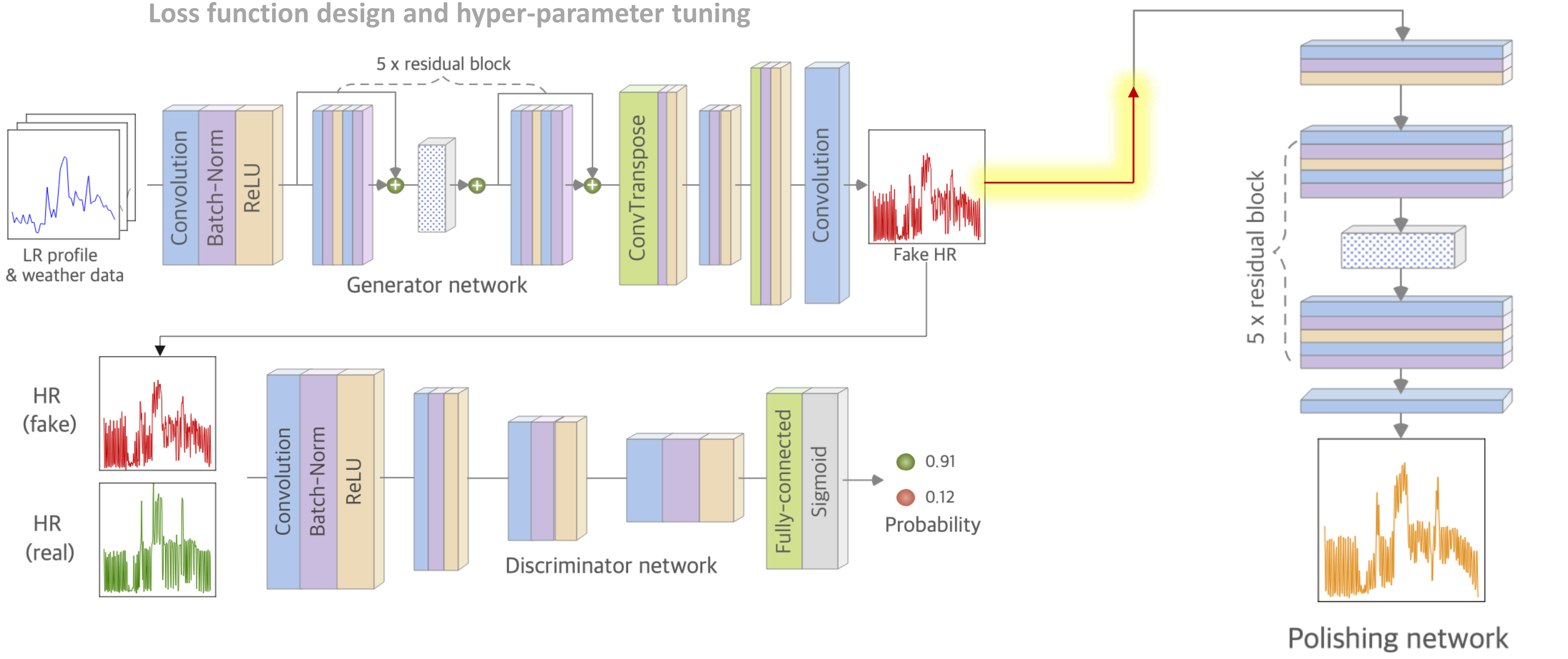
Lidong Song, Yiyan Li, and Ning Lu. "ProfileSR-GAN: A GAN based Super-Resolution Method for Generating High-Resolution Load Profiles." *IEEE Transactions on Smart Grid* 13, no. 4 (2022): 3278-3289. [Youtube video].

Stage 1: Inspired by the image processing applications

Loss function design and hyper-parameter tuning

Stage 2: fine-tuning

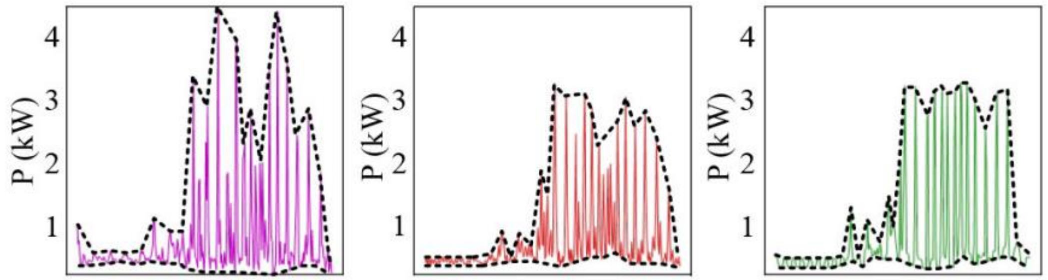
Power system domain expertise



Lidong Song, **Yiyan Li** and N. Lu, "ProfileSR-GAN: A GAN Based Super-Resolution Method for Generating High-Resolution Load Profiles," in *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3278-3289, July 2022, doi: 10.1109/TSG.2022.3158235. ProfileSR-GAN: https://www.youtube.com/watch?v=nBkwTqHplh8&t=30s

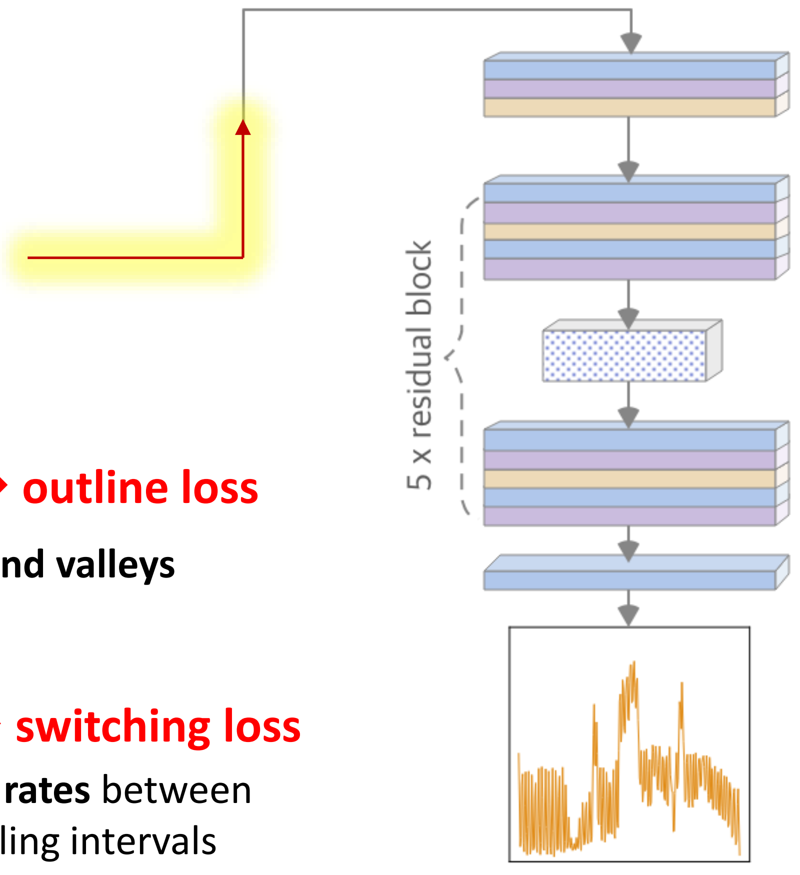
Stage 2: fine-tuning

use power system domain expertise



(a) Before polishing (b) After polishing (c) Ground truth

Fig. 7. An illustration of comparing the envelopes of the generated daily HR profiles (before and after polishing) with that of the actual daily load profile.



$$L_{pol} = L_{outl} + L_{swit} \tag{12}$$

$$L_{outl} = \frac{1}{N}\left\|\xi_{\max}\left(\hat{P}^{\mathrm{HR}}\right) - \xi_{\max}\left(P^{\mathrm{HR}}\right)\right\|_2^2 + \frac{1}{N}\left\|\xi_{\max}\left(-\hat{P}^{\mathrm{HR}}\right) - \xi_{\max}\left(-P^{\mathrm{HR}}\right)\right\|_2^2 \tag{13}$$

Shape Characteristics → outline loss

Compare **local peaks and valleys**

$$L_{swit} = \frac{1}{N}\left\|\xi_{\max}\left|\Delta\hat{P}^{\mathrm{HR}}\right| - \xi_{\max}\left|\Delta P^{\mathrm{HR}}\right|\right\|_2^2$$

$$\Delta\hat{P}^{\mathrm{HR}} = \hat{P}^{\mathrm{HR}}(n+1) - \hat{P}^{\mathrm{HR}}(n),$$

$$\Delta P^{\mathrm{HR}} = P^{\mathrm{HR}}(n+1) - P^{\mathrm{HR}}(n) \tag{14}$$

Ramp Characteristics → switching loss

Compare **load change rates** between two consecutive sampling intervals

Lidong Song, **Yiyan Li** and N. Lu, "ProfileSR-GAN: A GAN Based Super-Resolution Method for Generating High-Resolution Load Profiles," in *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3278-3289, July 2022, doi: 10.1109/TSG.2022.3158235. ProfileSR-GAN: https://www.youtube.com/watch?v=nBkwTqHplh8&t=30s

2. Generated from Scratch

Group load profile generation using GAN

Load Profile Generation Methods

TABLE I
COMPARISON OF OUR MULTILOAD-GAN MODEL WITH STATE-OF-THE-ART METHODS

		Description	Advantages	Disadvantages	Model output
Model-based methods [1][2]		Use physical models, such as building thermodynamics and customer behavioral models, to simulate electricity consumption profiles.	Explainable as the models reflect the laws of physics when describing the behavior behind field measurements	Require detailed physics-based models with many inputs and require parameter tuning.	**Single load profile** (When generating a load profile, the methods do not consider the spatial-temporal correlations among a group of generated load profiles)
Data-driven methods	Clustering based [3][4]	Cluster existing load profiles into different categories so that by combining the load profiles across different categories, SLPs are generated.	Easy to implement and can represent some realistic load profile characteristics.	Lack of diversity when using combinations of a limited number of existing profiles.	
	Forecasting based [5]-[8]	Generate SLPs based on publicly available load or weather data.	Easy to implement and flexible to generate load profiles with different lengths and granularities.	Depend heavily on historical data. The generated load profiles have similar patterns with historical data, therefore, lack of diversity.	
	SingleLoad-GAN-based [10]-[12] (the benchmark method)	GAN-based generative methods to generate the SLP for one customer at a time.	Learn from the real data distribution to generate diversified load profiles with high-frequency details.	Hard to train.	
	MultiLoad-GAN (the proposed method)	GAN-based generative methods to generate a group of spatial-temporal correlated load profiles simultaneously. Such load profiles can be loads served by the same transformer or feeder.	Learn from the distribution of real data to generate diversified load profiles with high-frequency details. Preserve the spatial-temporal correlations between loads.	Hard to train.	**Multiple spatial-temporal correlated load profiles**

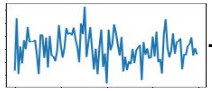
Yi Hu, Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "MultiLoad-GAN: A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167

Single-Load GAN Approach

Step 1: Generate one load profile at a time

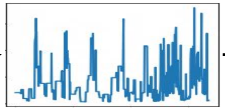
Step 2: Run step 1 iteratively to obtain a group of load profiles

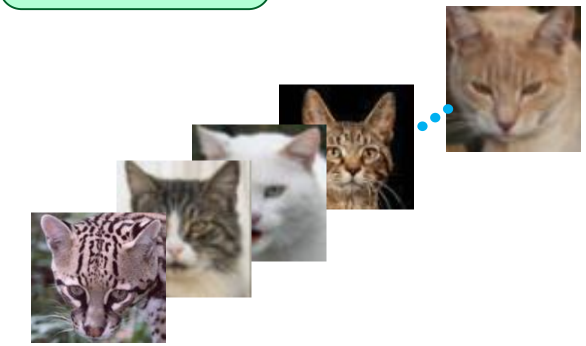
Noise



SingleLoad -GAN

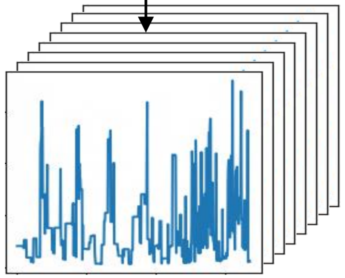
1 load profile





Load profile Database

Step 3: Randomly sample N load profiles to form a group of loads



8 load profiles

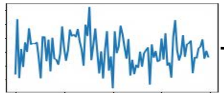
Drawbacks:
Cannot account for group-level characteristics

Yi Hu, Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "MultiLoad-GAN: A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167

Group-Load GAN Approach

Single-Load GAN

Step 1: Generate one load profile at a time



Noise

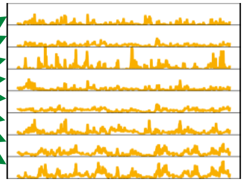
SingleLoad GAN

Step 2: Run step 1 for many times to obtain a database of load profiles

A load profile

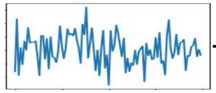
Load profile Database

Step 3: Randomly sample N load profiles



Capture group correlation

Group-Load GAN

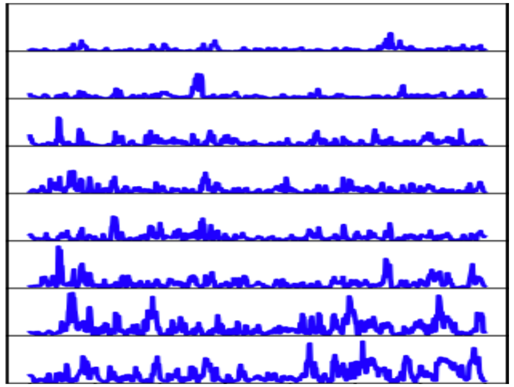


Noise

MultiLoad GAN

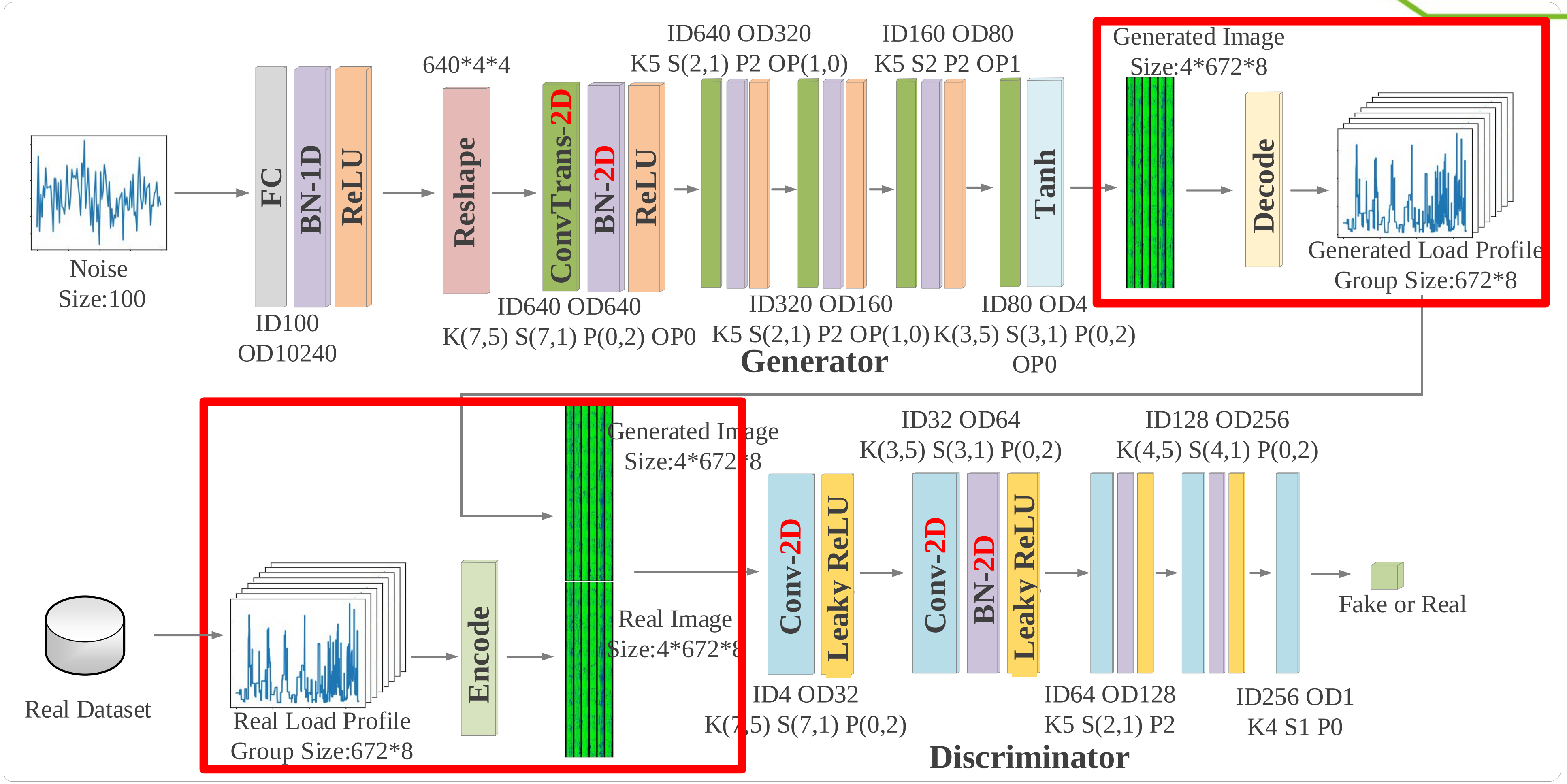
Generate N load profiles

A group of load profiles supplied by the same distribution transformer



Yi Hu, Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "MultiLoad-GAN: A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167

Configuration of MultiLoad-GAN

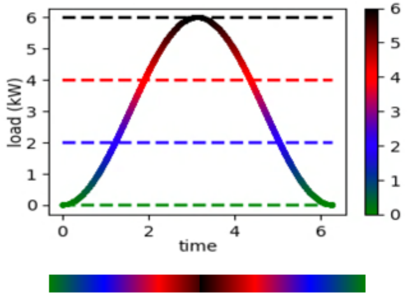


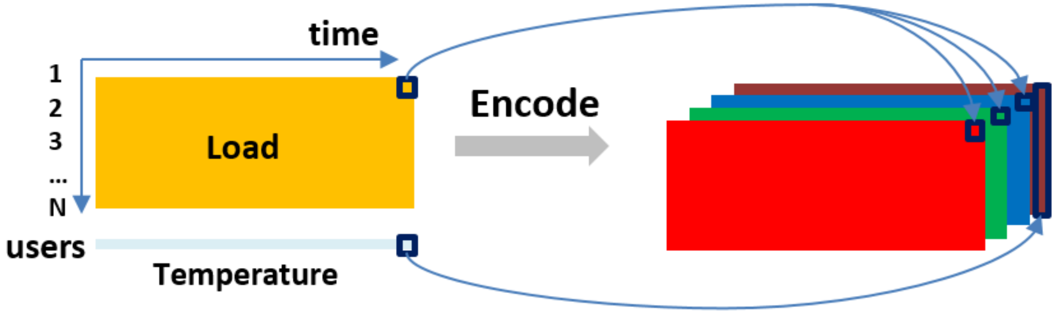
Profile-to-image encoding **2D-convolution layers**

Yi Hu, Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "MultiLoad-GAN: A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167

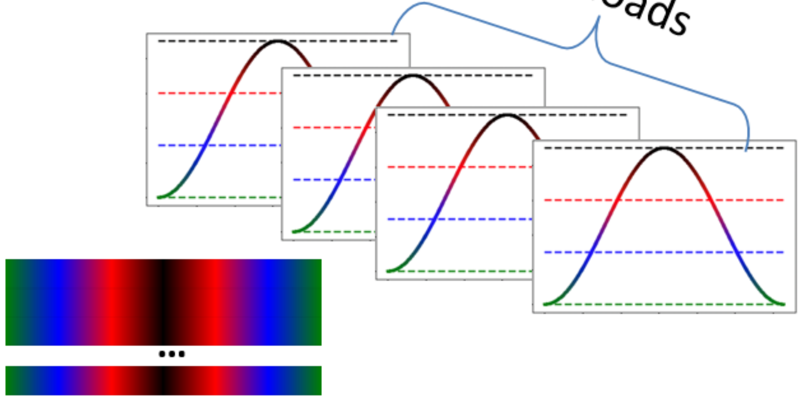
Profile-to-image Mapping

Profile-to-image Encoding: time-series plots to 4-channel ([r, g, b, t]) image





(a) Map one load profile to an image bar



(b) Map a group of loads to an image with N bars

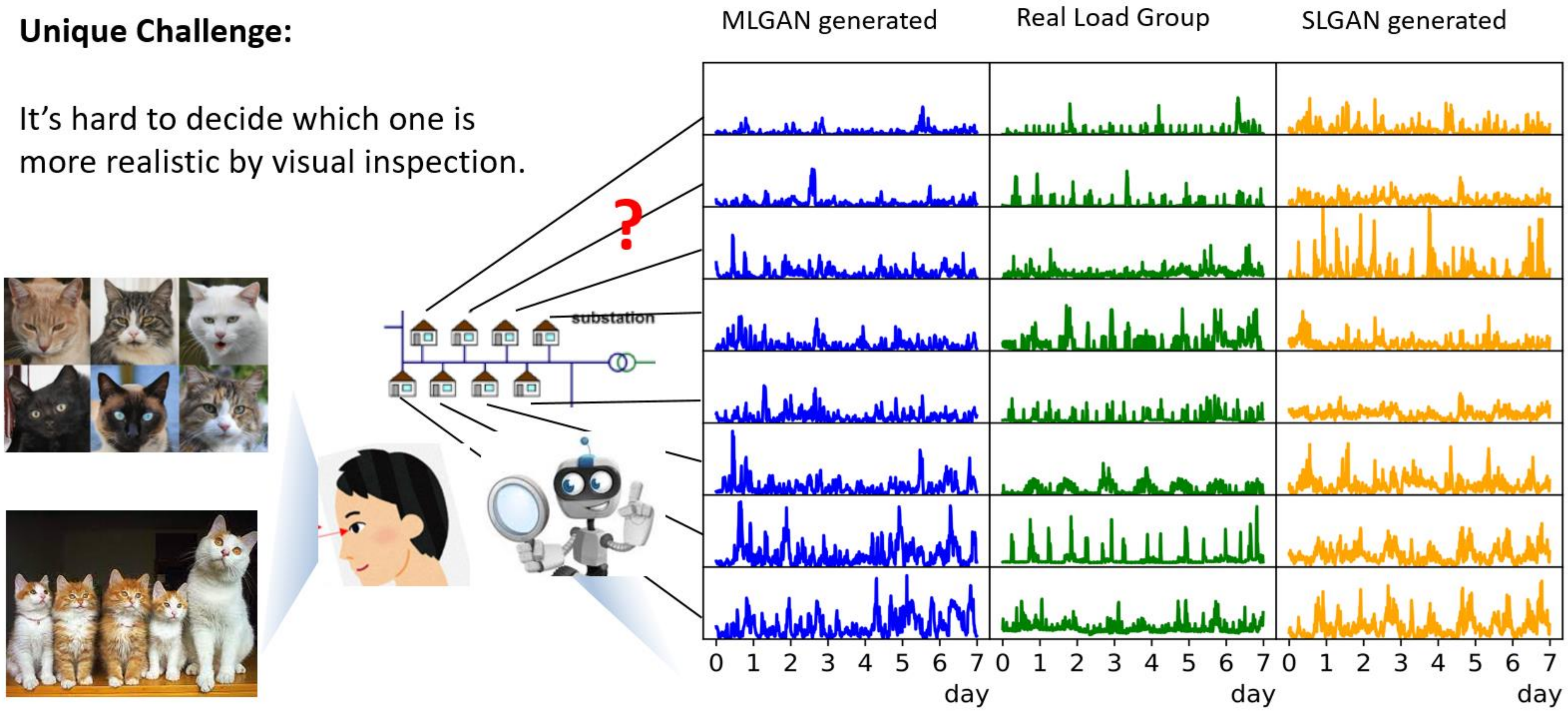
Load (kW)	[r, g, b]	Temperature(°F)	Vector [t]
0	[0, 1, 0]	0	[0]
(0, 2)	g↓, b↑		
2	[0, 0, 1]		
(2, 4)	b↓, r↑	(0,120)	t↑
4	[1, 0, 0]		
(4, 6)	r↓		
[6, +∞)	[0, 0, 0]	120	[1]

Yi Hu, Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "MultiLoad-GAN: A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167

How to Evaluate Realisticness?

Unique Challenge:

It's hard to decide which one is more realistic by visual inspection.



Yi Hu, Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "MultiLoad-GAN: A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167

Realisticness Evaluation Metrics

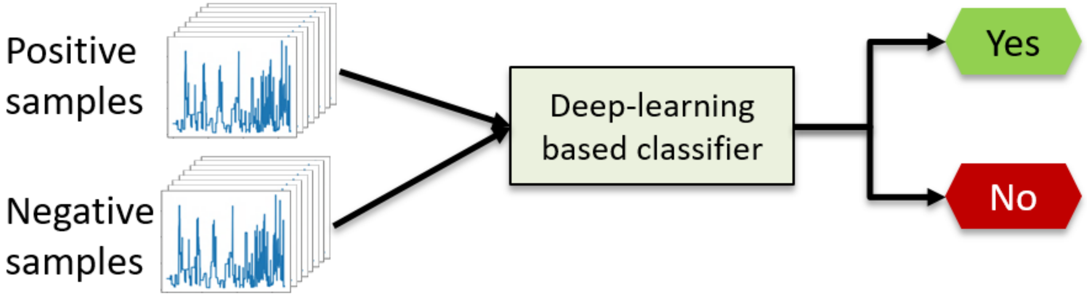
Statistical Evaluation

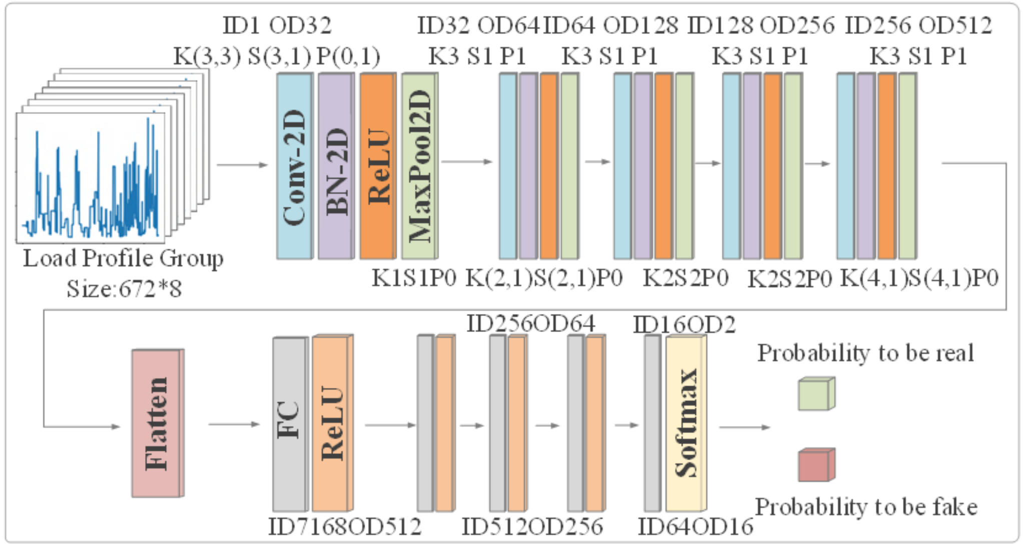
Whether or not group-level correlations are preserved?

Level	Indices
Household	Peak load distribution
	Mean power consumption distribution
	Load ramps distribution
	Hourly energy consumption distribution
	Daily energy consumption distribution
Transformer Level	Peak load distribution
	Mean power consumption distribution
	Load ramps distribution
	Hourly energy consumption distribution
	Daily energy consumption distribution

Deep-learning based Specialized Classifier

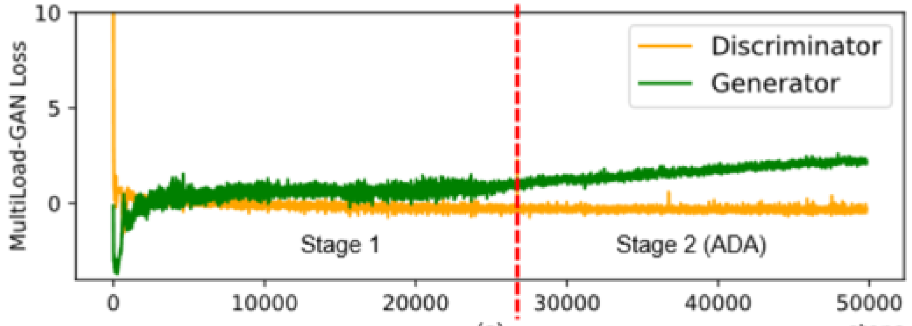
Whether or not high-level hidden features are similar?



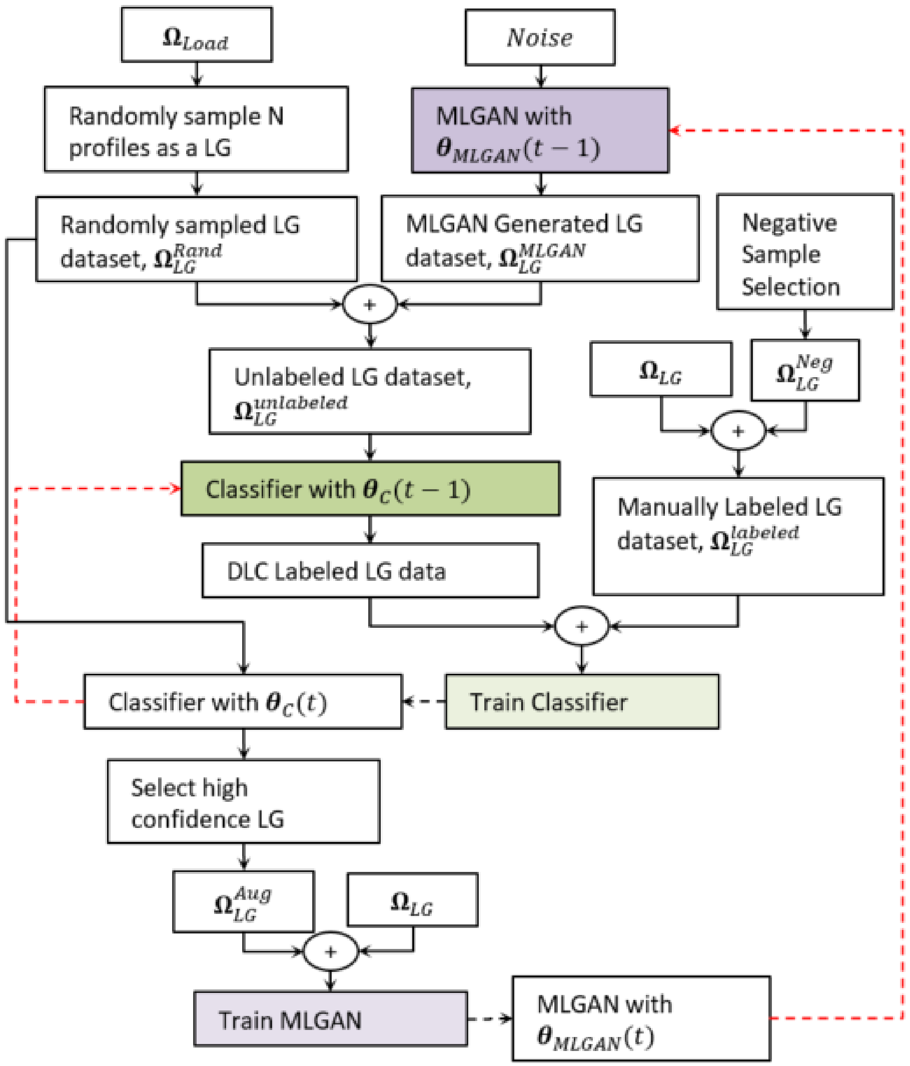


Iteratively Co-train GAN and Classifier

- We train the Classifier and MultiLoad-GAN iteratively.
- Then, let the partially trained classifier and MultiLoad-GAN generate augmented training data to enrich the training data set.
- This will improve the performance of both.



(a)



Prevent Over-train and Mode Collapse

1. Percentage of True

$$POR = \frac{Q_{real}}{Q} \times 100\%$$

2. Mean Confidence Level

$$MCL = \frac{1}{Q}\sum_{i=1}^{Q} P_{true}(i)$$

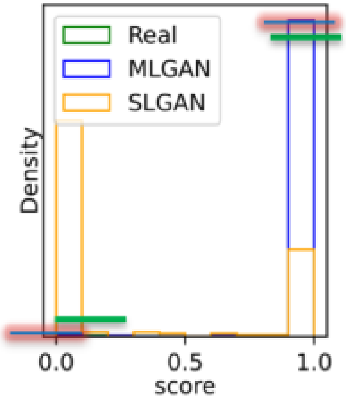
3. Confidence distribution

$$\tau(C(\boldsymbol{\Omega}_{LG})) = \tau([P_{true}(1), P_{true}(2), \dots, P_{true}(Q)])$$

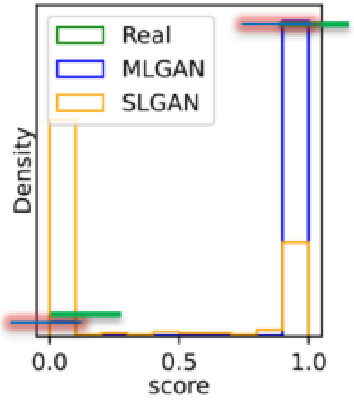
4. Freshet inception distance

$$Similarity = FID(\tau(\boldsymbol{\Omega}_{\mathbf{LG}}), \tau(\boldsymbol{\Omega}_{\mathbf{LG}}^{\mathbf{M}\mathit{L}\mathbf{GAN}}))$$

Dataset	Indices	Original	ADA Boosted
$\boldsymbol{\Omega}_{\mathbf{LG}}$	POR	94.38%	
	MCL	0.9371	
$\boldsymbol{\Omega}_{\mathbf{LG}}^{\mathbf{S}\mathit{L}\mathbf{GAN}}$	POR	19.69%	
	MCL	0.1913	
	FID with $\boldsymbol{\Omega}_{\mathbf{LG}}$	0.5173	
$\boldsymbol{\Omega}_{\mathbf{LG}}^{\mathbf{M}\mathit{L}\mathbf{GAN}}$	POR	99.06%	94.99%
	MCL	0.9899	0.9491
	FID with $\boldsymbol{\Omega}_{\mathbf{LG}}$	0.01106	**0.000055**



W/O ADA



With ADA

Conclusions

- **Future test systems should be digital-twin based**
  - Enable a virtual playground for researchers and developers to develop new grid support functions
  - Compared with field tests, testing on digital twins are safer, cheaper, faster, and scalable
  - The key to digital-twin based power system models lies in synthetic data and topology generation.

- **Challenges**
  - A substantial collection of realistic network topologies and high-resolution data sets is needed
  - Encompass extensive geographical areas and utilities.
  - Standardized validation process and comprehensive sets of evaluation criteria are needed.
  - High-quality publicly available data sets play a crucial role in benchmarking various generative algorithms, enabling performance comparisons, and driving advancements in synthetic data generation technology.

References

1. **Yi Hu,** Yiyan Li, Lidong Song, Han Pyo Lee, PJ Rehm, Matthew Makdad, Edmond Miller, and Ning Lu, "**MultiLoad-GAN:** A GAN-Based Synthetic Load Group Generation Method Considering Spatial-Temporal Correlations," submitted to IEEE Transactions on Smart Grid (2022). Available online at: https://arxiv.org/abs/2210.01167
2. **Lidong Song, Yiyan Li,** and Ning Lu. "**ProfileSR-GAN: A** GAN based Super-Resolution Method for Generating High-Resolution Load Profiles," http://arxiv.org/abs/2107.09523, Youtube video.
3. **Ming Liang**, Y. Meng, J. Wang, D. Lubkeman and N. Lu, "**FeederGAN:** Synthetic Feeder Generation via Deep Graph Adversarial Nets," in IEEE Transactions on Smart Grid, doi: 10.1109/TSG.2020.3025259.
4. **Kai Ye,** Hyeonjin Kim, Di Wu, PJ Rehm, and Ning Lu, "A Modified Sequence-to-point HVAC Load Disaggregation Algorithm," submitted to 2023 IEEE PES General Meeting, Available online at: https://arxiv.org/abs/2212.04886. 23PESGM1248
5. **Hyeonjin Kim,** Kai Ye, Han Pyo Lee, Rongxing Hu, Di Wu, PJ Rehm, and Ning LU, "An ICA-Based HVAC **Load Disaggregation** Method Using Smart Meter Data" submitted to 2023 ISGT. Available online at: https://arxiv.org/abs/2209.09165
6. **Wang, Jiyu,** Xiangqi Zhu, Ming Liang, Yao Meng, Andrew Kling, David L. Lubkeman, and Ning Lu. "A Data-Driven Pivot-Point-Based Time-Series Feeder **Load Disaggregation** Method." IEEE Transactions on Smart Grid 11, no. 6 (2020): 5396-5406.
7. **Ming Liang,** Jiyu Wang, Yao Meng, Ning LU, David Lubkeman, and Andrew Kling. "A Sequential **Energy Disaggregation** Method using Low-resolution Smart Meter Data, " Proc. of IEEE Innovative Smart Grid Technologies, Washington DC, 2019.