

SDCC Liaison Meeting

CSI

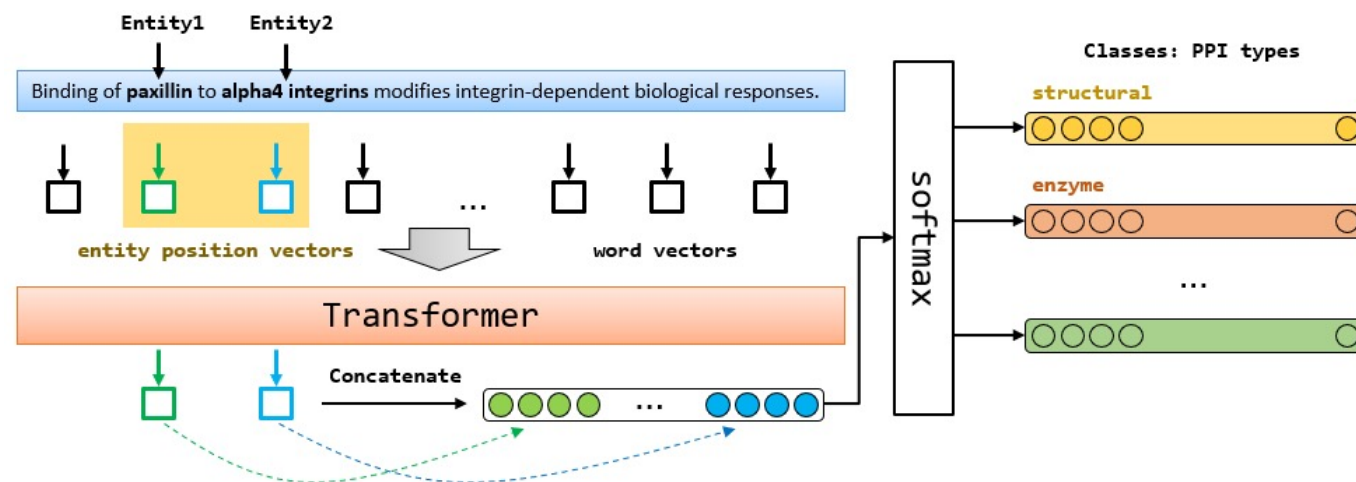
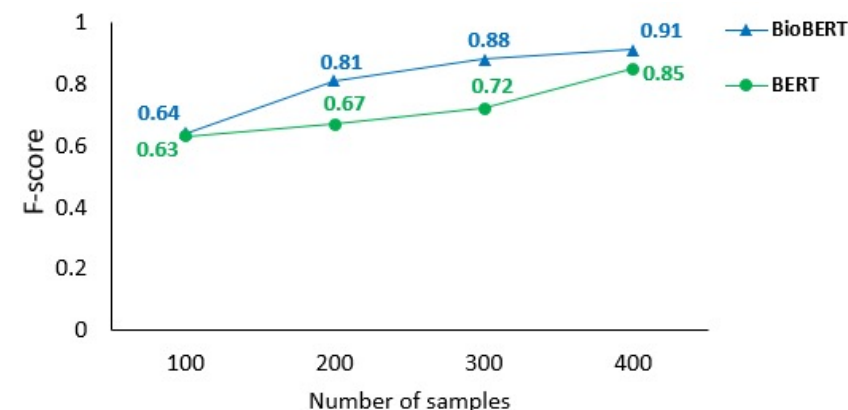
Yihui Ren
yren@bnl.gov

Volta Partition

- Combat Covid (NLP and GNN)
- STLearn (AI for multivariable temporal learning)
- SciDAC-RAPIDS (climate science)
- SciDAC-HEP (cosmology)
- Advanced Data Acquisition LDRD (sphenix)

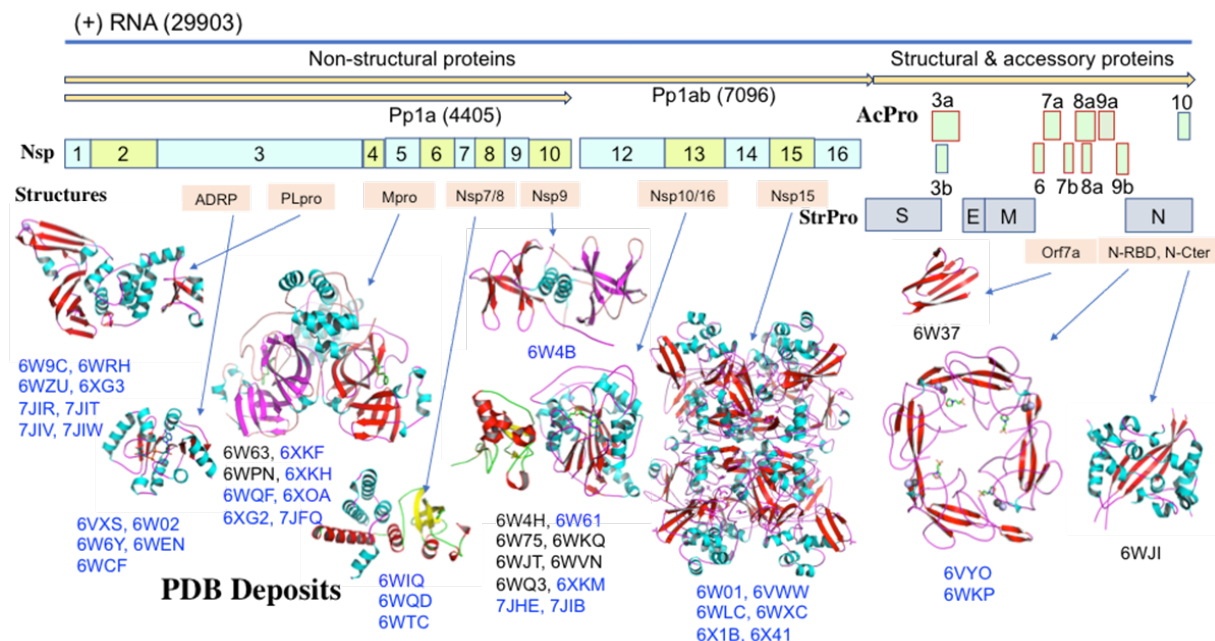
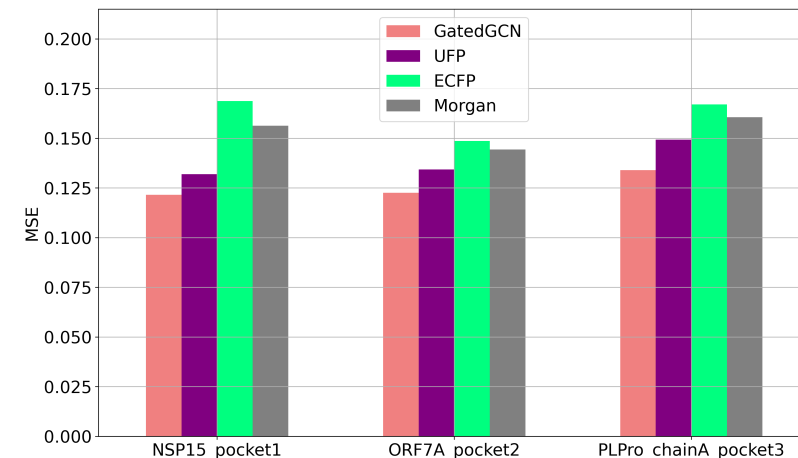
NLP: Named Entities Relation Learning from Literature

- Relational information is a foundation for knowledge base construction.
- Multi-stage learning is error prone.
- Joint learning in identifying named entities and their biological relationship (i.e., protein-protein interaction)



GraphNN: Molecular Fingerprints

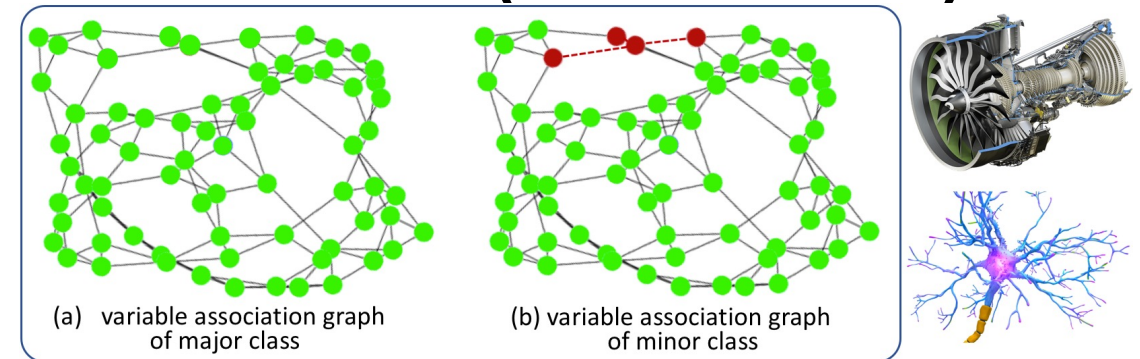
- Trained on over 300K docking compounds and 18 covid-19 related protein sites.
- Fast and Learnable molecular fingerprinting method.
- Perform better than traditional methods.



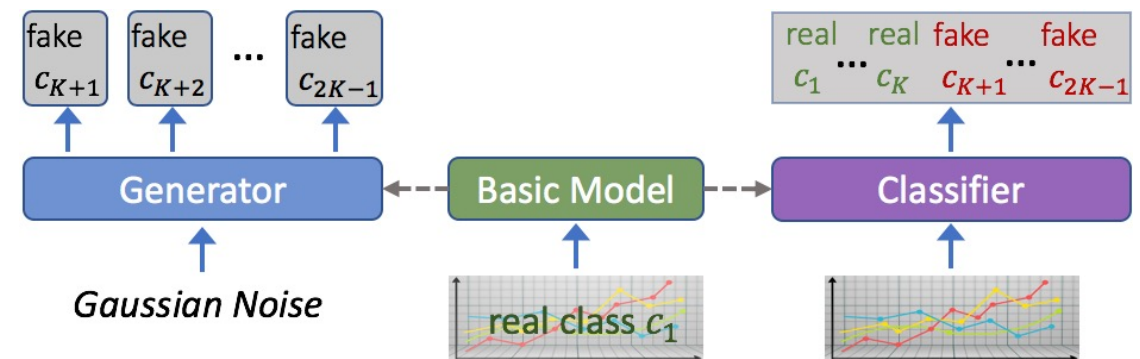
Target	Baseline	ECFP	Morgan	GatedGCN	GraphSAGE	MPNN
3CLPro_pocket1	1.830	1.132	1.115	1.031	1.096	1.061
ADRP-ADPR_pocket1	1.588	0.237	0.202	0.151	0.157	0.151
ADRP-ADPR_pocket5	1.584	0.237	0.200	0.149	0.156	0.153
ADRP_pocket1	0.544	0.115	0.109	0.085	0.092	0.087
ADRP_pocket12	0.544	0.116	0.109	0.085	0.091	0.086
ADRP_pocket13	1.048	0.163	0.144	0.104	0.111	0.104
COV_pocket1	0.270	0.076	0.070	0.054	0.058	0.058
COV_pocket2	0.271	0.075	0.069	0.055	0.058	0.057
COV_pocket8	0.872	0.178	0.162	0.125	0.133	0.127
COV_pocket10	1.166	0.172	0.163	0.120	0.124	0.124
NSP9_pocket2	1.139	0.205	0.211	0.164	0.170	0.167
NSP9_pocket7	0.987	0.135	0.126	0.089	0.093	0.092
NSP15_pocket1	0.843	0.169	0.156	0.122	0.126	0.122
ORF7A_pocket2	0.397	0.149	0.144	0.123	0.127	0.124
PLPro_chainA_pocket3	0.566	0.167	0.161	0.134	0.142	0.135
PLPro_chainA_pocket23	0.927	0.253	0.242	0.199	0.207	0.195
PLPro_pocket6	0.843	0.157	0.139	0.112	0.118	0.110
PLPro_pocket50	1.335	0.293	0.272	0.211	0.222	0.215

STLearning: Interpretable Temporal Generative Adversarial Networks (IT-GANs)

- **Classifier:** IT-GANs is a time series data classification algorithm with imbalance class distribution.
 - GAN is a powerful tool for such imbalance class distribution training.
- **Interpretable:** Identifies the difference among classes by learning an association graph for each class.
- **Generative:** IT can simulate system behavior using a generative model, which is helpful, especially for minor classes.

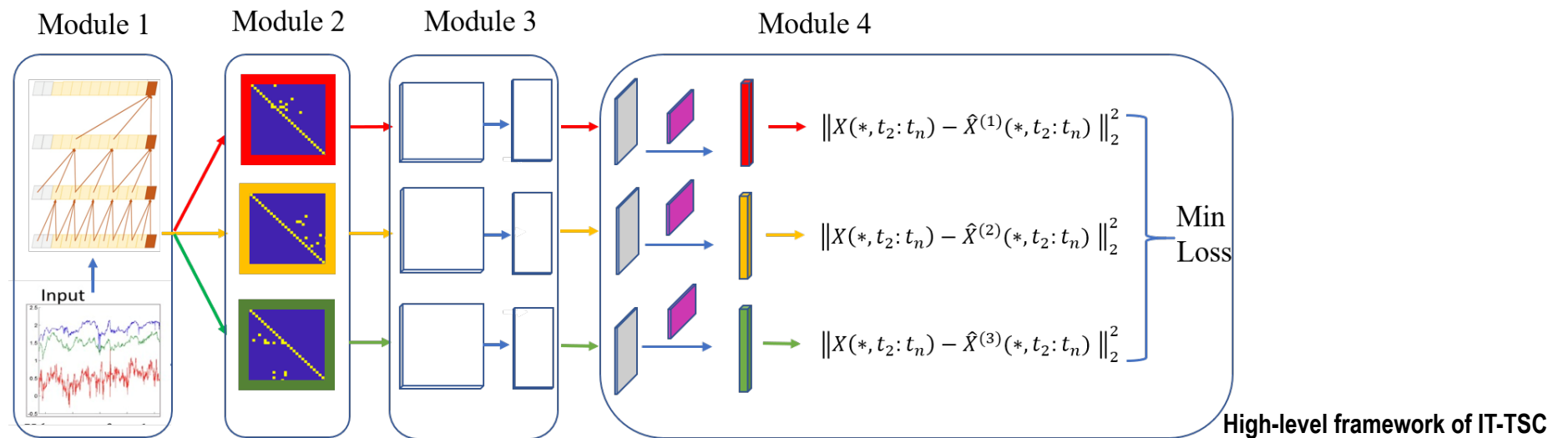


In variable association graphs. Each node represents a variable in a complex industrial system, and the edges indicate the association between two connected nodes. The edges in red indicate the association change in a minority class (right) compared with the majority class (left). Learning these unknown associations is crucial for understanding the behavior of system, but it also is challenging due to data imbalance.



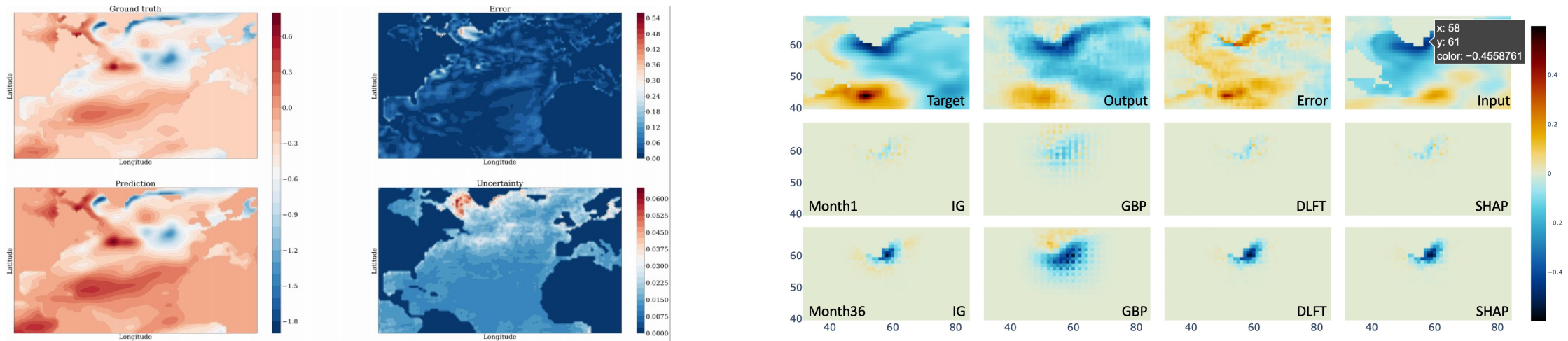
STLearning: Interpretable Time Series Clustering (IT-TSC)

- Clustering and each cluster variable association graphs are explored together.
- Time series autoregression by a multi-path neural network, where each path corresponds to one cluster.
- The learned variable association graphs can be used to interpret how one cluster differs from another.



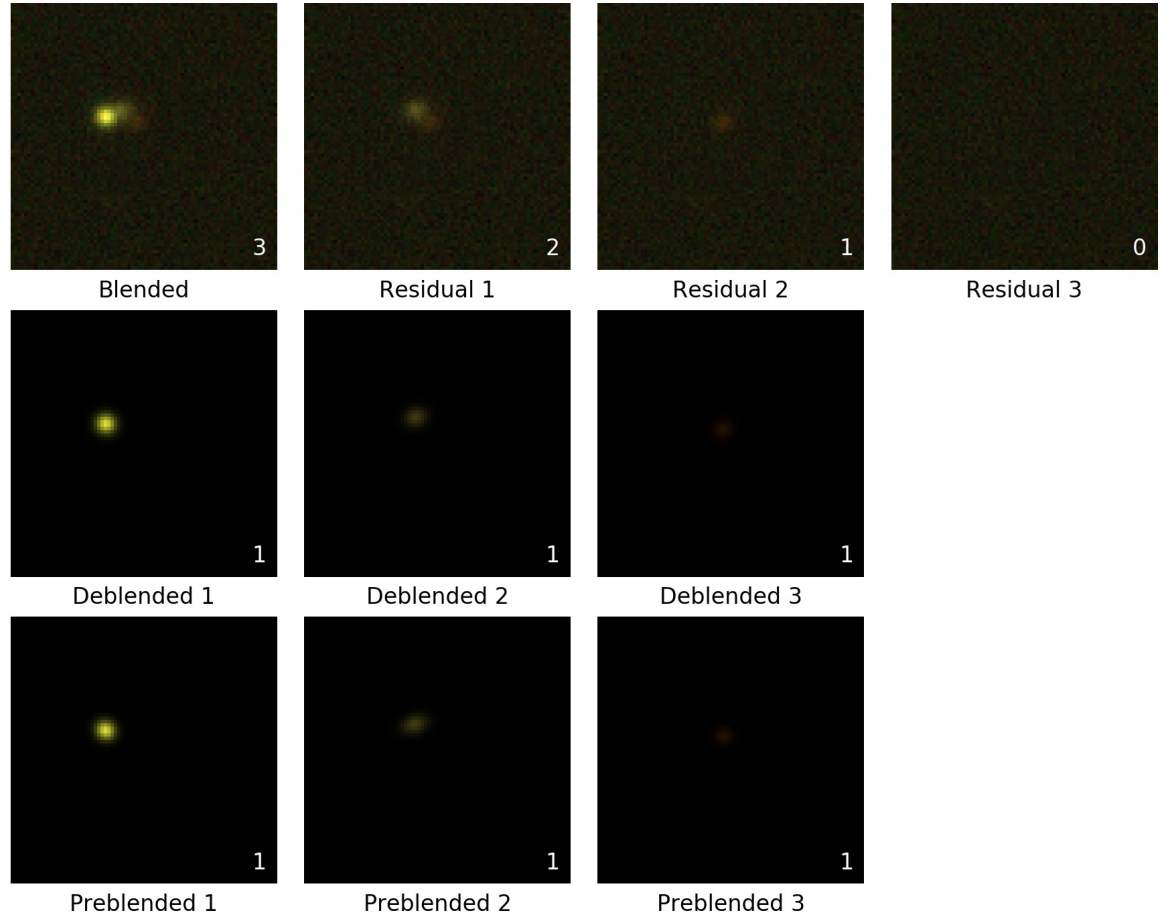
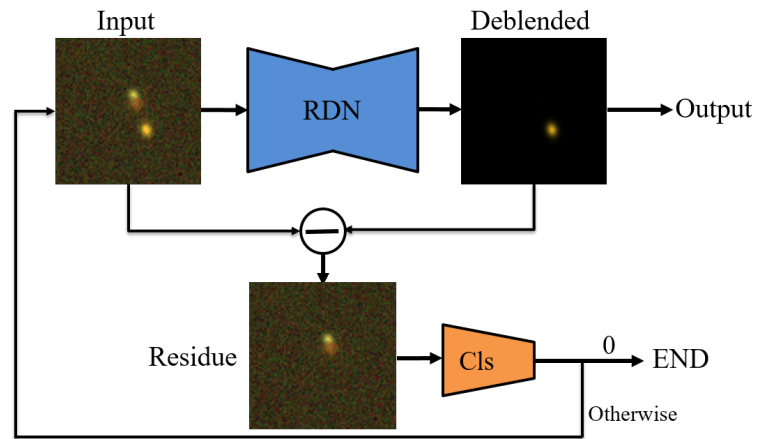
SciDAC-RAPIDS: Bayesian Neural Network and xAI on Climate Prediction

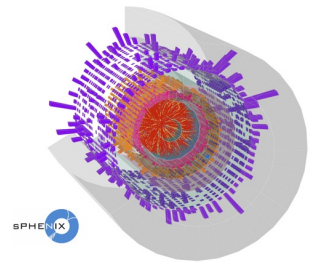
- Applied various xAI approach to understand temporal dependency behaviors.
- Discovered local interaction only and working on improving the model to capture longer-term dependencies.
- Bayesian Neural Network is also used to quantify uncertainties and also improved forecasting accuracies compared to non-Bayesian approach.



SciDAC-HEP: Iterative Galaxy Deblending

- Iteratively deblending one galaxy at a time.
- Deep residual dense network was used for one galaxy deblending.
- Classifier predicts how many galaxies are present and subtracts deblended galaxy if any are left.

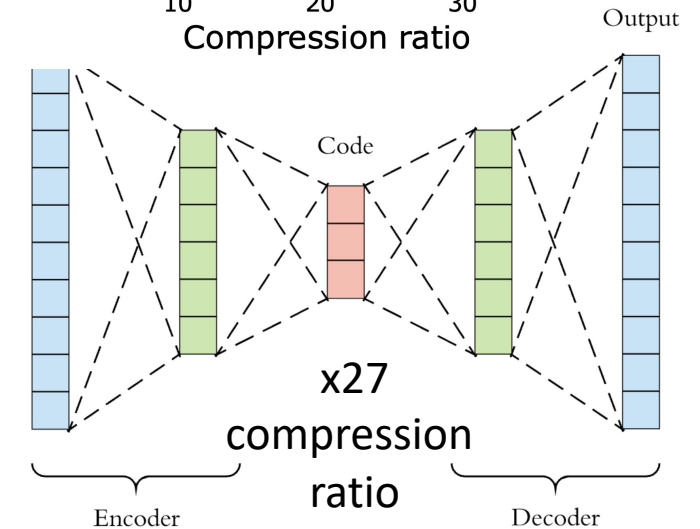
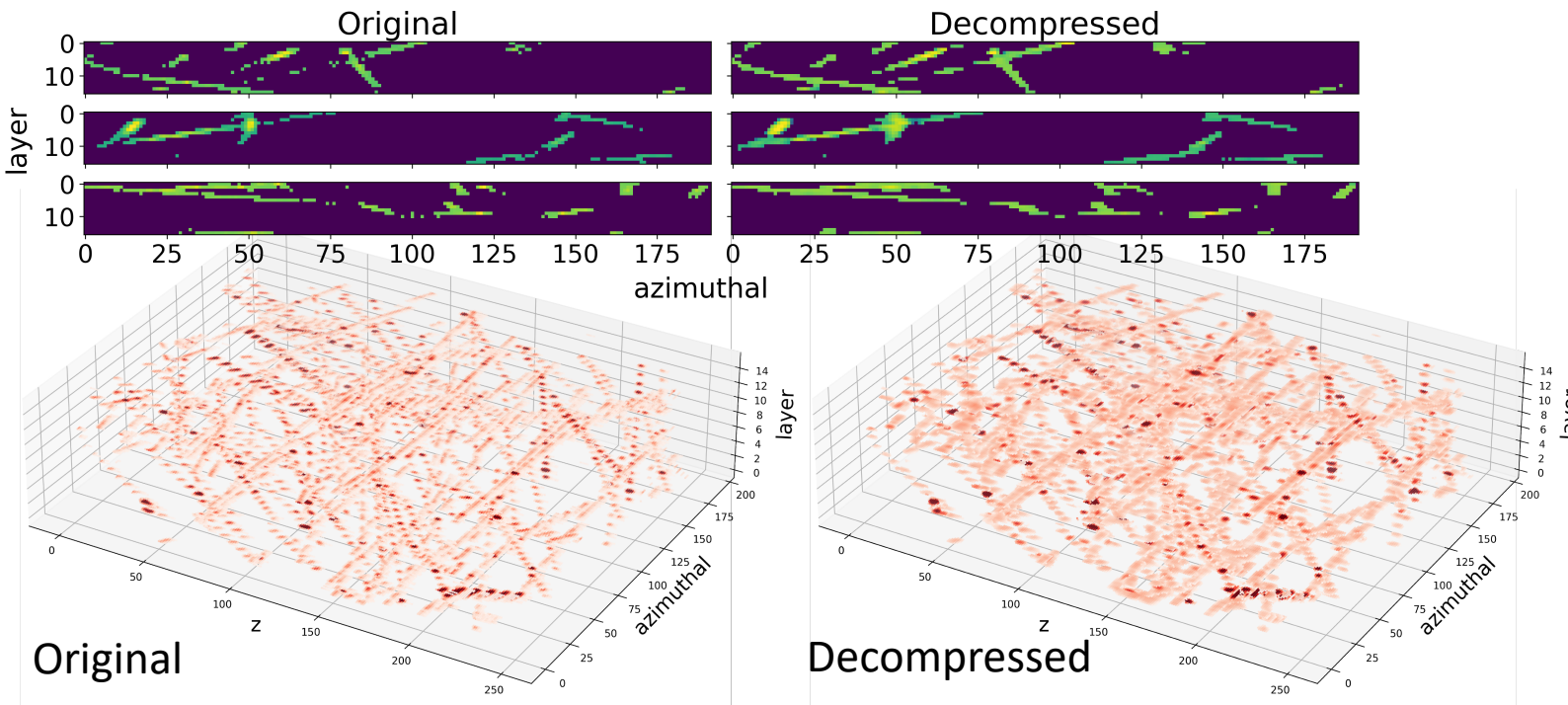
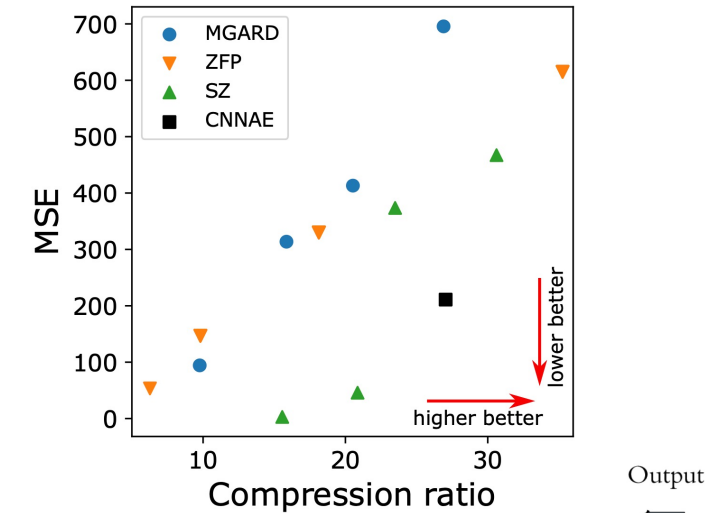




Advanced Data Acquisition LDRD

- 3D Sparse 10-bit TPC data.
- CNN auto-encoder with binary pixel-wise classifier.
- Offers better compression ratio and lower MSE

Comparing with other lossy compression algorithms



Thanks!

- Extremely fast response
- JupyterHub
- DockerHub