ML @ CSI

BROOKHAVEN
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Machine Learning Improves Solar Energy Forecasting

Scientific Achievement

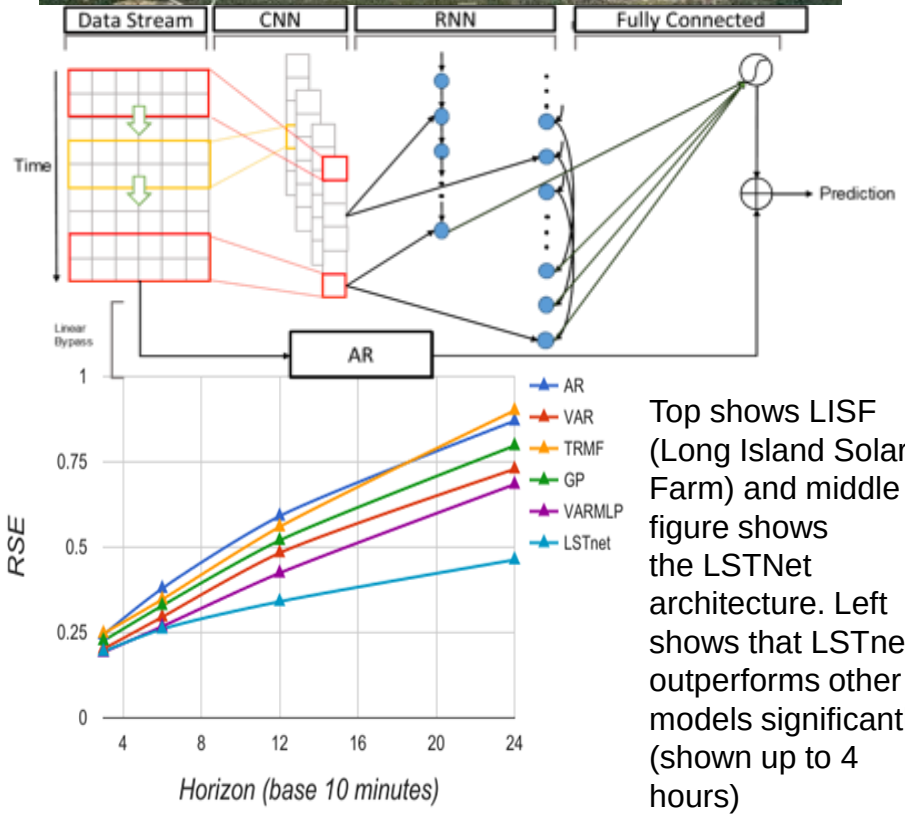
Our newly developed spatio-temporal learning algorithm (LSTnet) significantly improves solar energy forecasting

Significance and Impact

Utility scale spatio-temporal solar irradiance forecasting algorithms can improve operation and integration of solar farms with electric power grid and they can be applied to any spatio-temporal data

Research Details

- Joint modeling of short and long-term time dependency by combining Deep Learning (DL) and Autoregressive (AR) models
- CNN (Convolutional Neural Network) captures local dependency patterns whereas RNN (Recurrent Neural Network) captures long-term dependency patterns
- Automatically adapting the mixtures weights of the AR and CNN/RNN components based on input



Top shows LISF (Long Island Solar Farm) and middle figure shows the LSTNet architecture. Left shows that LSTnet outperforms other models significantly (shown up to 4 hours)

ML increases AMI Load Forecasting Accuracy

Scientific Achievement

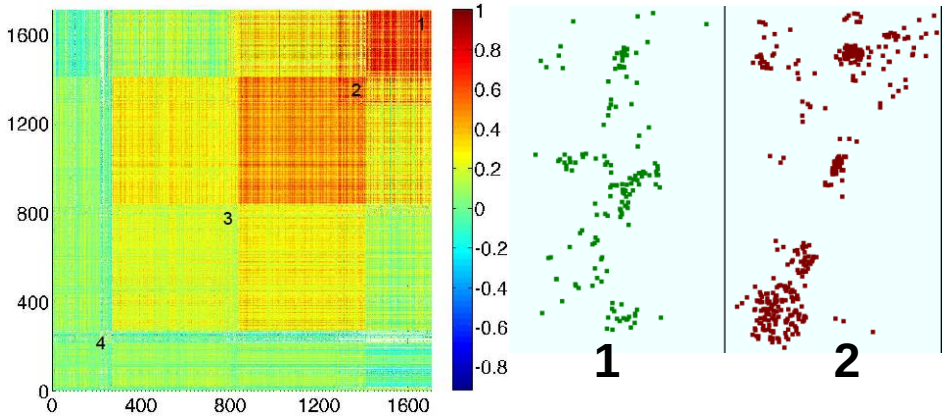
Individual household level load forecasting using Smart Meter

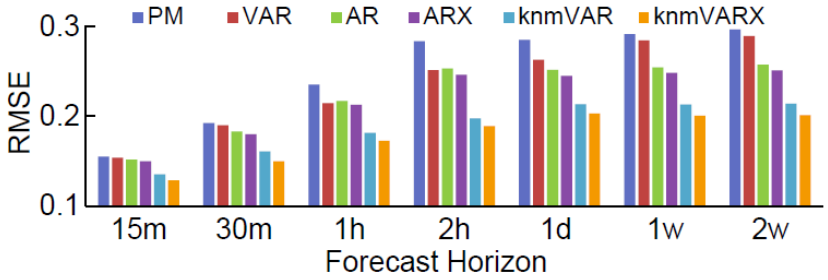
Significance and Impact

Utilizing spatio-temporal load patterns to significantly improve load forecasting accuracy and our algorithm can be easily applied to real-time distributed load forecasting

Research Details

- Constrained time series analysis using clustered customer pattern to capture spatio-temporal relationships better
- Weather and other factors are modeled as exogenous variables
- Applying AoW (Analysis on the Wire) framework to load forecasting enables real-time analytics while data is in transit





Top left shows four clustered Smart Meter customers using clustering analysis and top right illustrate the spatial distribution of cluster 1 and 2. Cluster 1 contains about 92% commercial customers and their spatial pattern is quite different from cluster 2 (80% residential customer). By exploiting this fact, we proposed knm (k nearest meter) based VARX (Vector AutoRegressive model with eXogenous variables), leading to significantly better forecasting accuracy

Machine Learning Emulator for Climate Modeling

Scientific Achievement

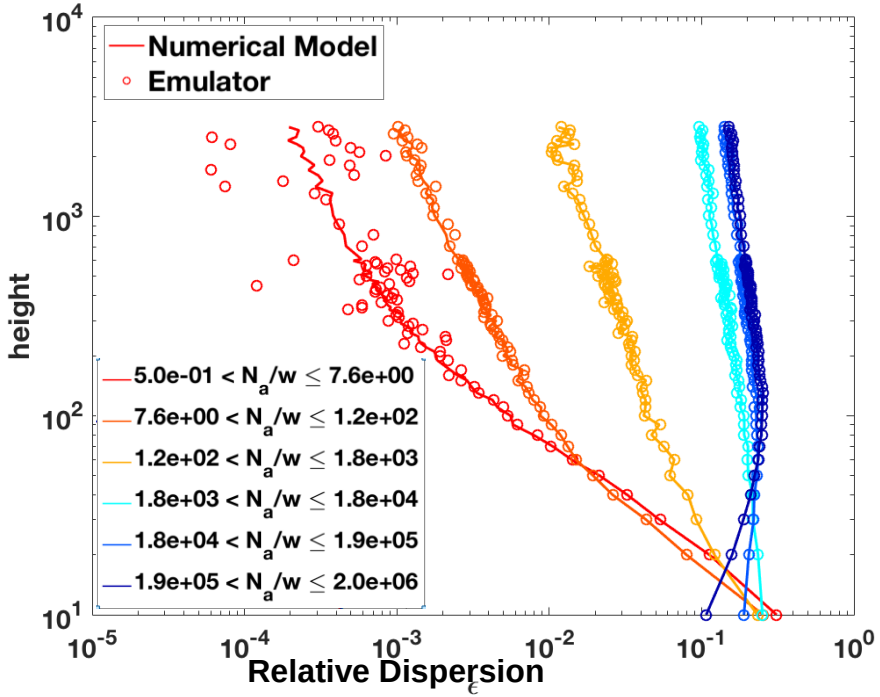
Accurately modeled Parcel model using temporal learning algorithm

Significance and Impact

Proof of concept in expensive physics based climate simulation with cheaper machine learning alternative

Research Details

- Successfully modeled Parcel models as a Gaussian Process (tested with other baseline methods)
- Created special monotonic transformation for significantly better Gaussian Process modeling accuracy
- Working on more general approaches and challenging science cases



The figure shows examples of the variations of key cloud properties (shown relative dispersion) with height simulated by a physics based numerical model and the ML emulator, under different environmental conditions characterized by the ratio of aerosol concentration **Na** and updraft velocity **w**. Other variables are also emulated well (not shown due to space limitations)

Streaming Unsupervised Feature Selection

Scientific Achievement

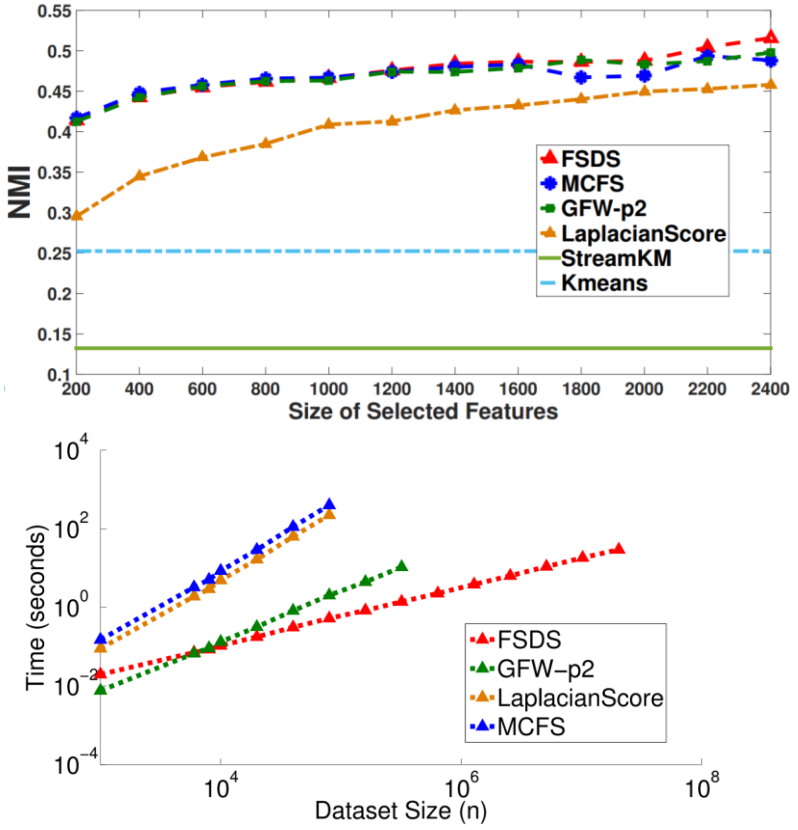
Streaming approximation (limited memory, one pass) of expensive manifold feature selection algorithm

Significance and Impact

Our proposed FSDF (Feature Selection on Data Stream) enables high velocity big data stream analysis and can be used with any machine learning algorithm as it is unsupervised

Research Details

- Matrix sketch based streaming decomposition with theoretical error bound analysis
- Instead of expensive Lasso regression, adopt cheaper Ridge regression on orthonormal space
- Dynamic feature subset selection as the topic or distribution changes



Compared to the state of the art batch unsupervised feature selection method (MCFS), our proposed FSDS (Feature Selection on Data Stream) showed similar or better accuracy on 20 newsgroup data (top) and yet much better scalability (bottom)

- Huang, Hao, Shinjae Yoo, and Shiva Prasad Kasiviswanathan. "Unsupervised feature selection on data streams." CIKM 2014.

Streaming Spectral Clustering (SSC)

Scientific Achievement

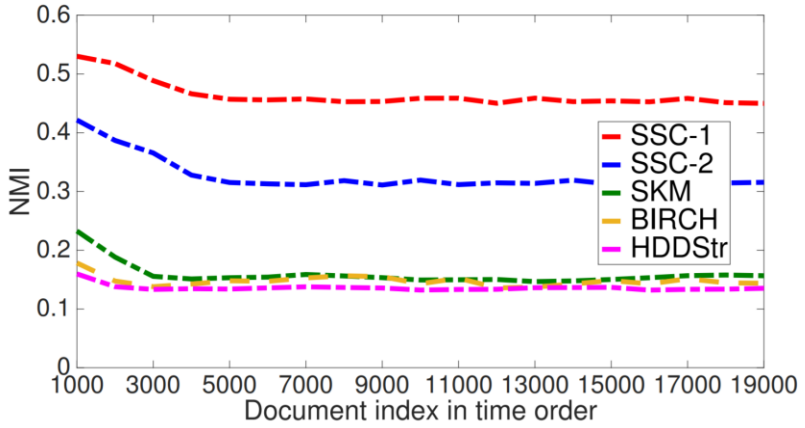
Streaming approximation (limited memory, one pass) of expensive spectral clustering algorithm

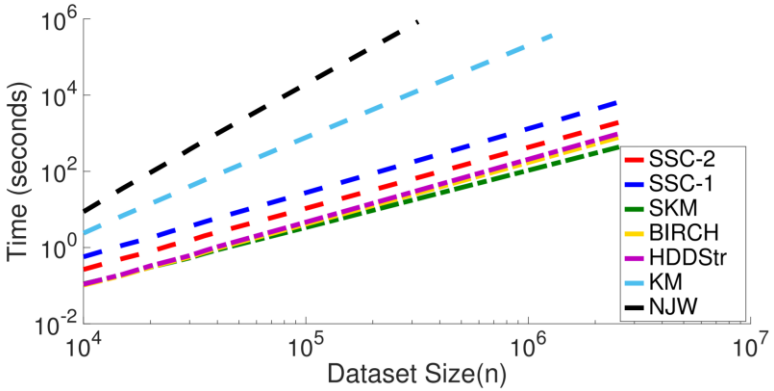
Significance and Impact

Our proposed SSC enables big data spectral clustering analysis without a big cluster. Anyone with a laptop can do high quality clustering analysis.

Research Details

- Streaming approximations of graph Laplacian matrix construction including degree matrix and symmetric normalization
- Extended matrix sketch algorithms for bigger stream batch size processing for better streaming decomposition approximation
- Adaptive to concept drift or distribution changes using the proposed manifold embedding rotation matrix



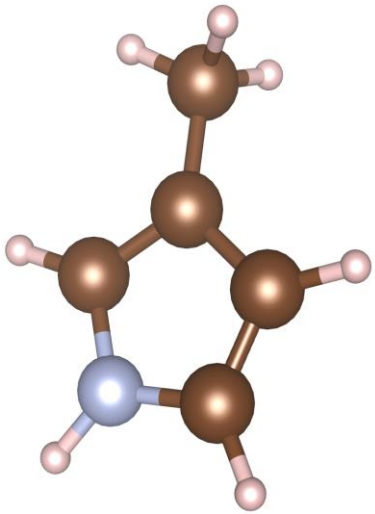


Compared to the state of the art streaming clustering algorithms, our proposed SSC showed much better accuracy (top) and yet comparable scalability (bottom)

Yoo, Shinjae, Hao Huang, and Shiva Prasad Kasiviswanathan. "Streaming spectral clustering." Data Engineering (ICDE), 2016 IEEE 32nd International Conference on. IEEE, 2016.

Forward/backward learning

ATOMIC STRUCTURE



experiment — NSLS-II

- *limited resource*
- *very high cost*
- *needs allocation*

simulation — FEFF

- *may be inaccurate*

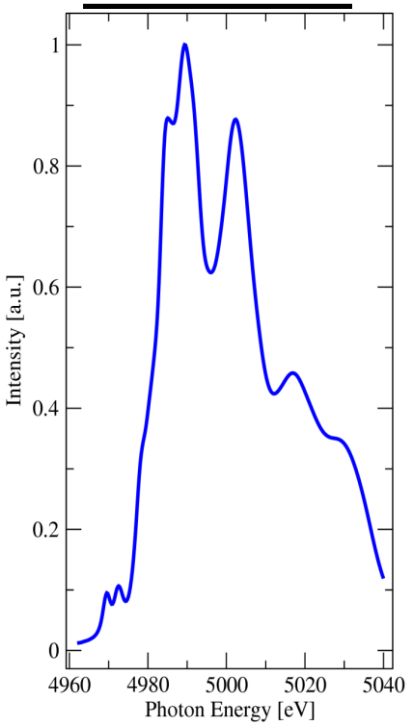
Forward Learning

ML

Backward Learning

ML may offer predictions in both directions

ABSORPTION SPECTRUM



Backward Learning – From spectra to Structure

Scientific Achievement

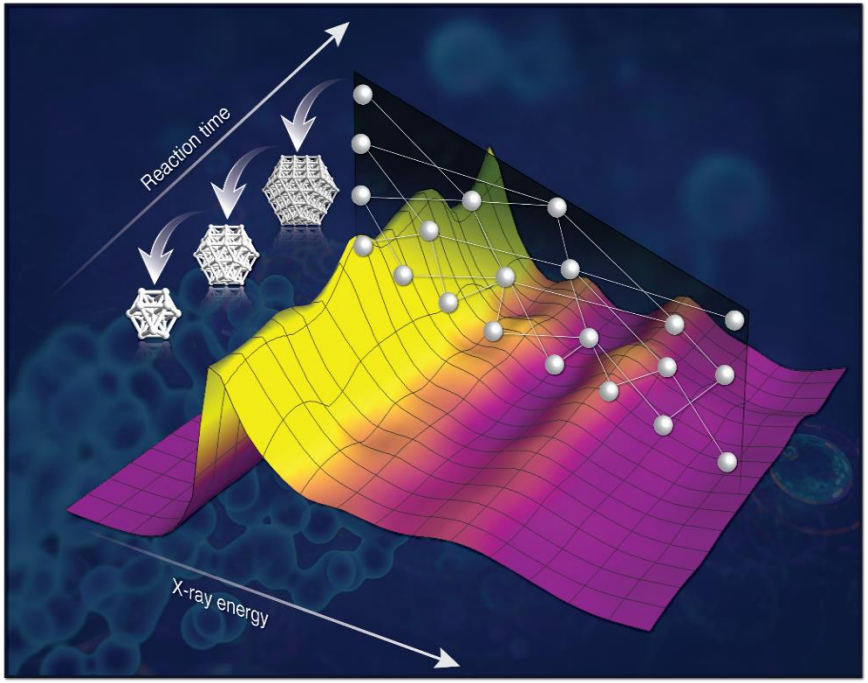
A method was developed to decipher the structure of nanoparticle catalysts from their X-ray absorption spectra.

Significance and Impact

Tracking the structure of catalysts in real working conditions is a challenge due to the paucity of experimental techniques that can measure the number of atoms and the distance between metal atoms with high precision. Accurate structural analysis at the nanoscale, now possible with the new method in the harsh conditions of high temperature and pressure, will enable new possibilities for tuning up reactivity of catalysts at the high flux and high energy resolution beamlines of NSLS-II and other advanced synchrotron facilities.

Research Details

Using machine learning techniques, the previously "hidden" relationships between the features in the X-ray absorption spectra and geometry of nanocatalysts were found.

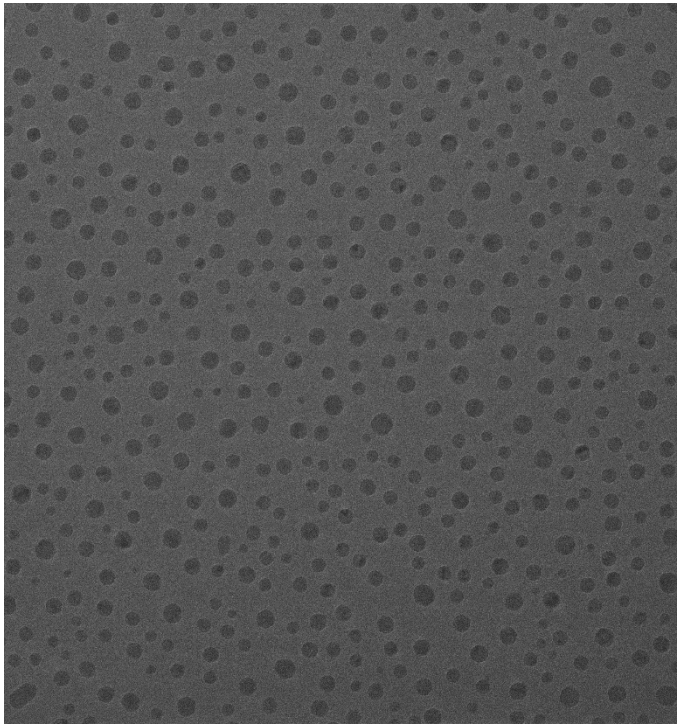


A sketch of the new method that enables fast, "on-the-fly" determination of three-dimensional structure of nanocatalysts. The colored curves are synchrotron X-ray absorption spectra collected in real time, during catalytic reaction (in operando). The neural network (white circles and lines) converts the spectra into geometric information (such as nanoparticle sizes and shapes) and the structural models are obtained for each spectrum.

J. Timoshenko, D. Lu, Y. Lin, A. I. Frenkel,
***J. Phys. Chem. Lett.*, 8 (20), 2017.**

Environmental Transmission Electron Microscope (ETEM)

ETEM



- Size: 1920×1792
- 400 frames/second
- 3GB/seconds
- High noise images

Nanoparticle detection in ETEM images

Scientific Achievement

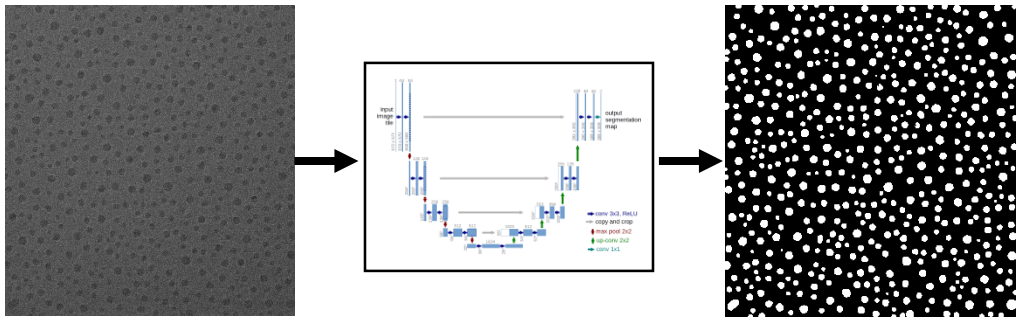
Developed a deep learning model to simultaneously detect and segment particles in ETEM images, without any labels needed

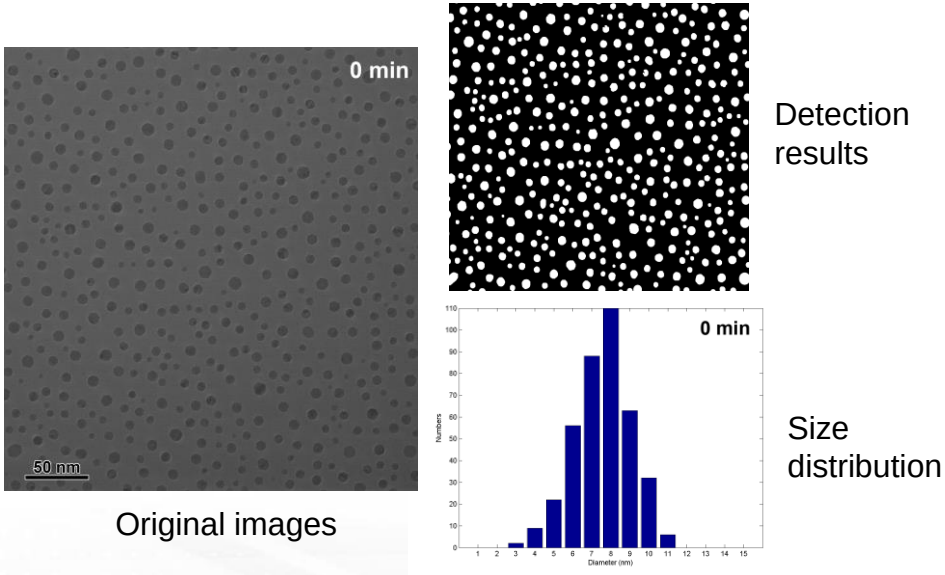
Significance and Impact

Finding nanoparticles in ETEM images accurately is the critical first step for automatically analyzing the physical characteristics of the nanoparticles.

Research Details

- A fully convolutional network (FCN) – U-Net is used for detection and segmentation simultaneously
- There is no manual labeling needed. It first applies traditional image segmentation techniques, and treats the result as the noisy label to literately train the U-Net.
- Once the U-Net is trained, it will be utilized to each frame independently, and it can be highly parallelized





Thank You!