Artificial intelligence activities

Arman Davtyan, Danilo Ferreira de Lima, Oleksii Turkot, ... Luca Gelisio
Data Analysis Group

Yue Sun, Christian Plückthun,... Péter Hegedűs, Sandor Brockhauser
Data Analysis Group, University of Szeged

Arman Davtyan: arman.davtyan@xfel.eu
Danilo Ferreira de Lima: danilo.enoque.ferreira.de.lima@xfel.eu
Yue Sun: yue.sun@xfel.eu

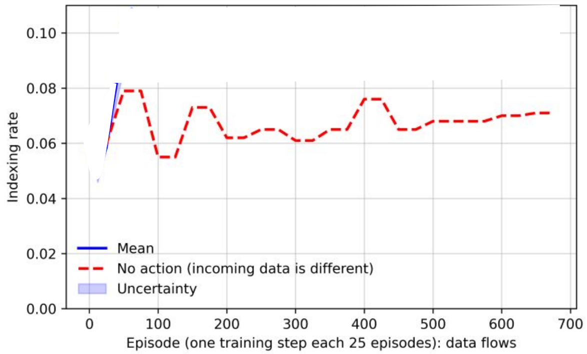
Hamburg, January 2022

Machine Learning applications for:

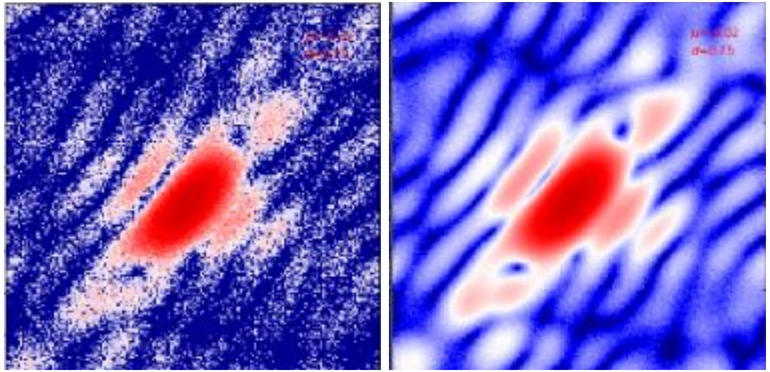
- ❖ Online analysis automation
- ❖ Data quality boosting
- ❖ Data clustering
- ❖ End-to-end ML methods for spectra classification

ML in action

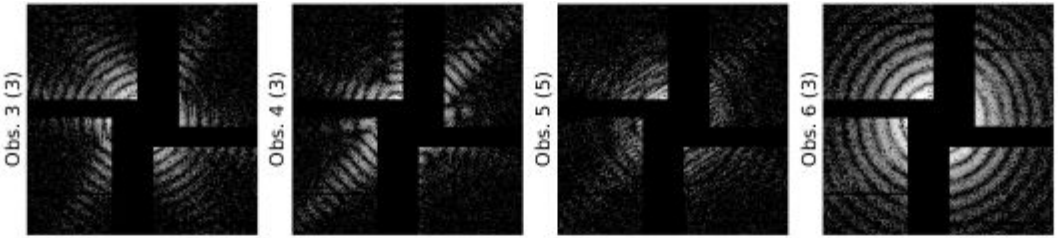
Task Automation

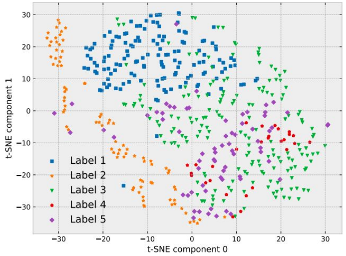


Data Quality Boosting



Data Clustering



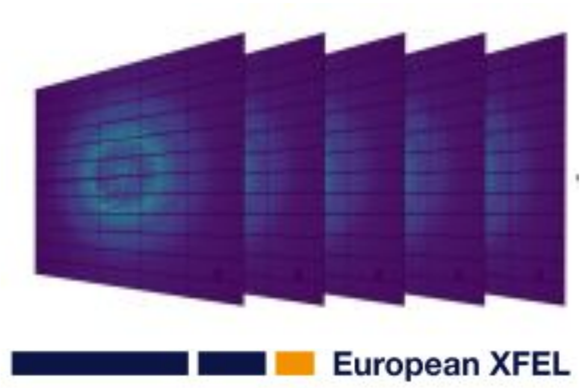


Reinforcement Learning for tuning CrystFEL parameters

CrystFEL

The CrystFEL software suite provides a data processing pipeline for serial femtosecond crystallography (SFX) experiments.

T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty and H. N. Chapman. “CrystFEL: a software suite for snapshot serial crystallography”. J. Appl. Cryst. 45 (2012), p335–341. doi:10.1107/S0021889812002312

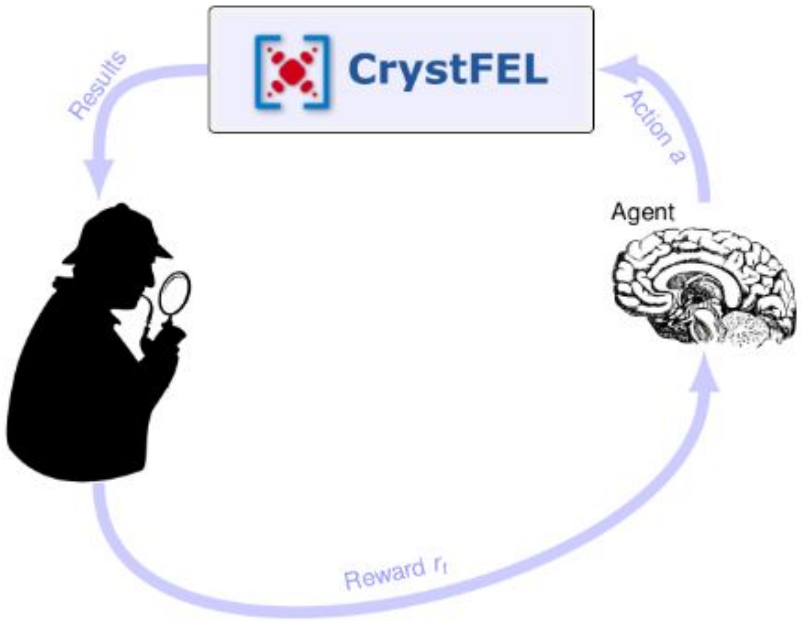


During a SFX experiment, one aims to collect enough indexable (i.e. resulting from a crystal) patterns.

Experimental variables such as the ones related to sample delivery are tuned so to optimize the indexing rate (the fraction of indexable patterns).

But CrystFEL also requires some parameters to be tuned in order to deliver a reliable indexing rate.

Train the agent using CrystFEL input and output



CrystFEL requires a number of parameters that have to be user adjusted.

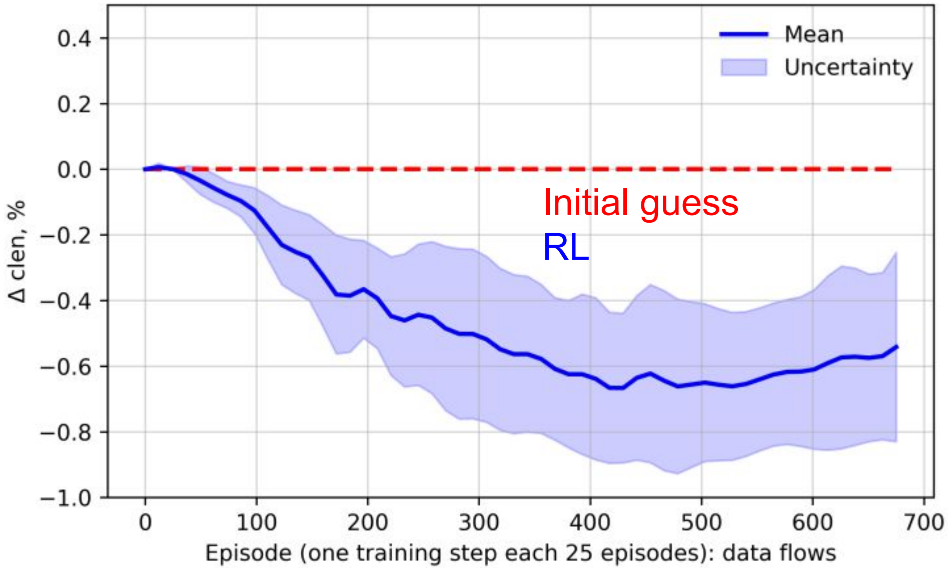
Reinforcement learning could assist the user to fine tune some parameters.

Reward = Δ (fraction of indexed patterns)

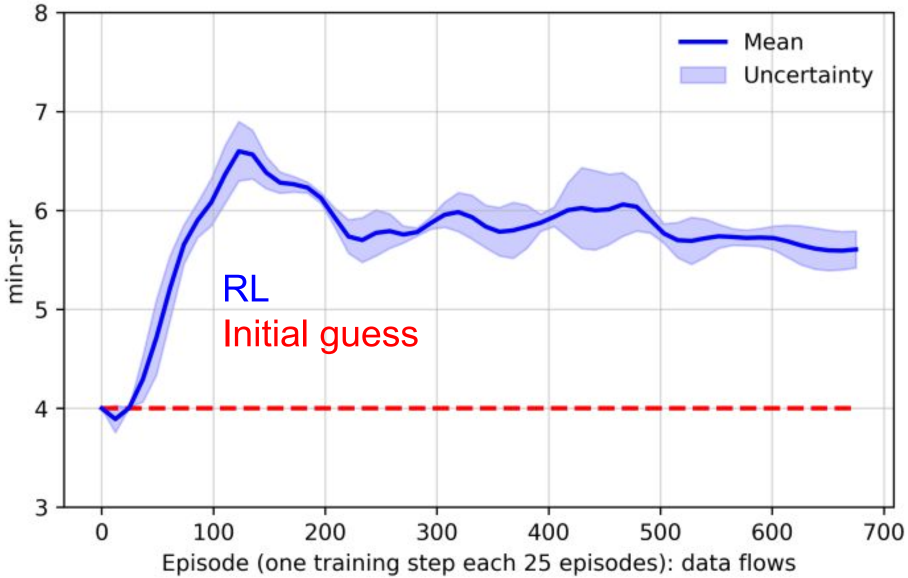
Reinforcement Learning Results

Parameters tuned: *min-snr, min-pix-count, clen*

param: *clen*



param: *min-snr*



European XFEL

Reinforcement Learning Results

- Optimize the indexing rate

- Assist to tune the initial guess

- Identify new use cases for RL application

Indexing rate

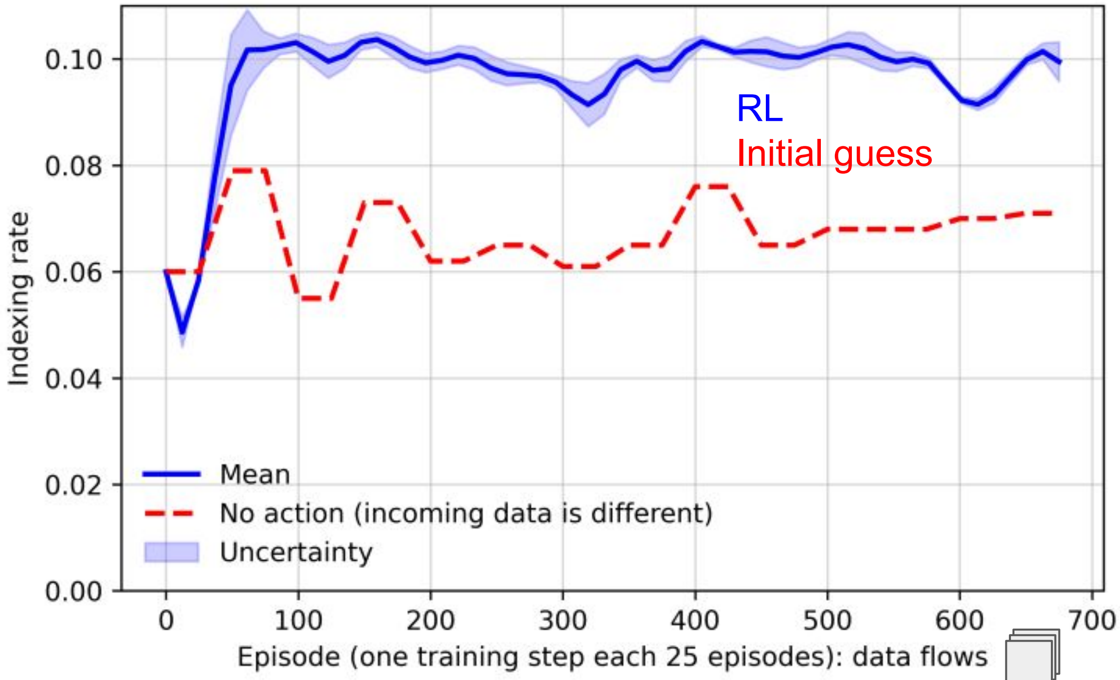
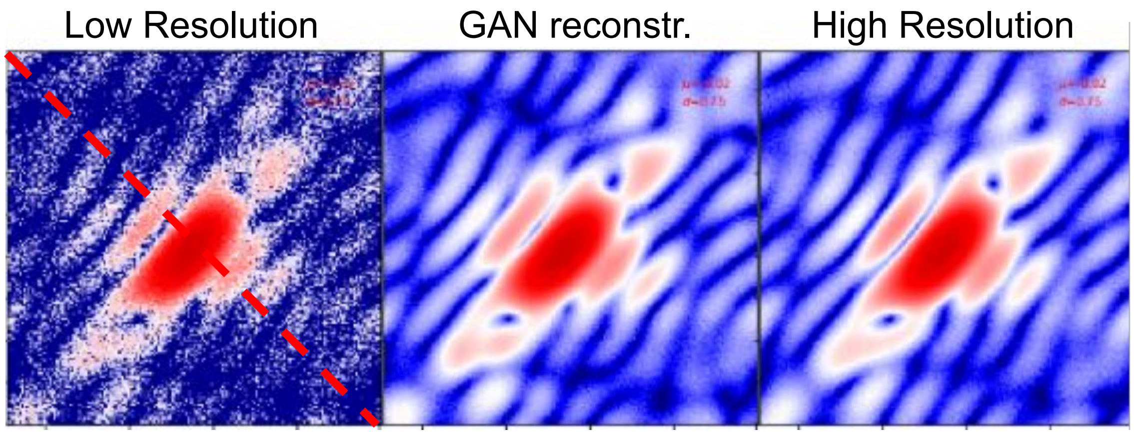
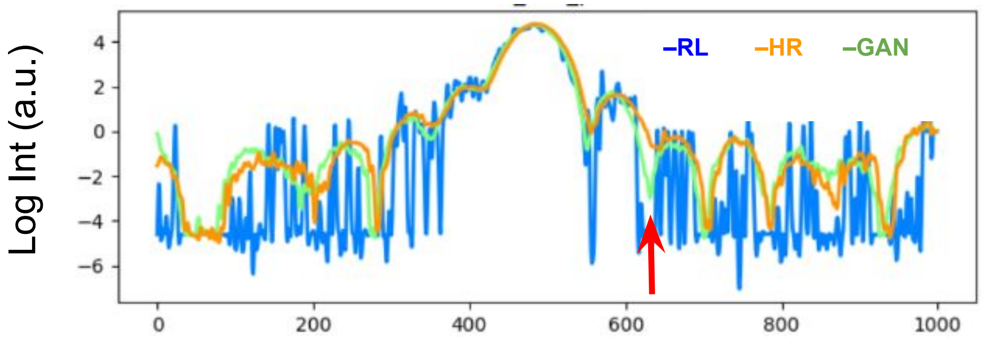


Image quality booster (GAN)

- Use Generative Adversarial Networks (GANs) to map image from low-resolution (LR) domain to high-resolution (HR) domain.
- LR and HR images *not paired*.
- Use test sample to learn the mapping function.
- Simulated data used
- Working on uncertainty …





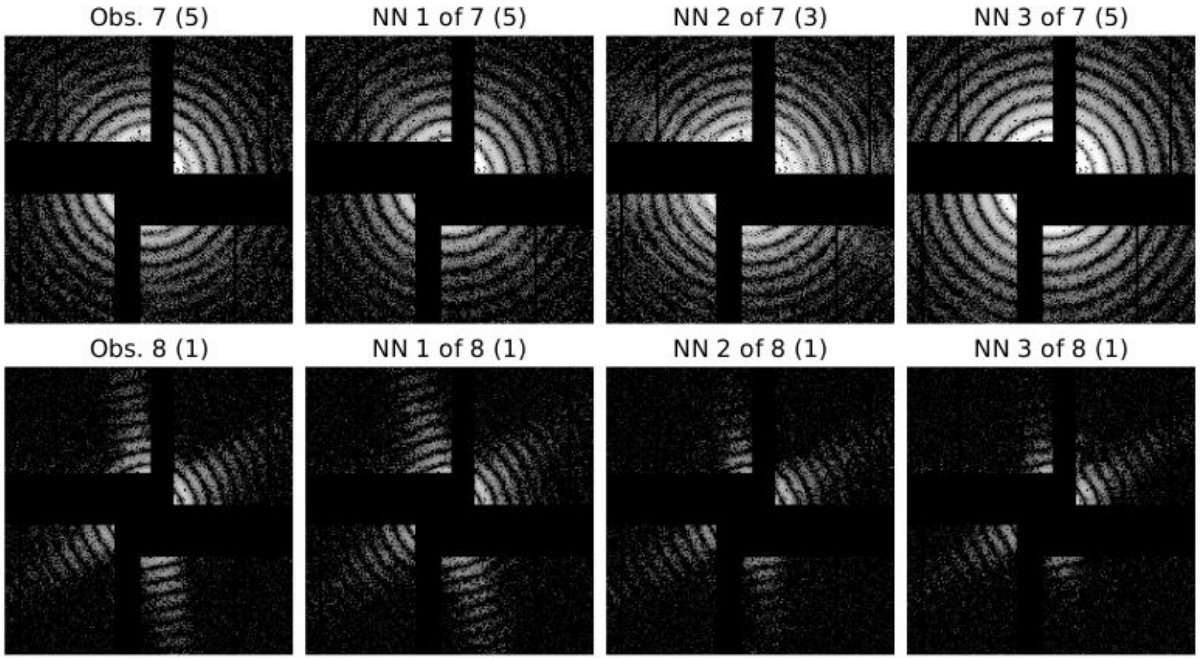
Line along the diagonal: LR, HR and GAN

Step 1: generate an augmentation-independent latent space

Step 2: classify latent space based on the nearest neighbours.

Clustering

- ❖ Find clusters in terabytes of data.
- ❖ Automatic experiment-independent clustering and labeling: single hits, multiple hits, no hits …
- ❖ Augmentation examples: Poisson noise, Gaussian noise, rotation …
- ❖ Method tested on simulated data (Condor)

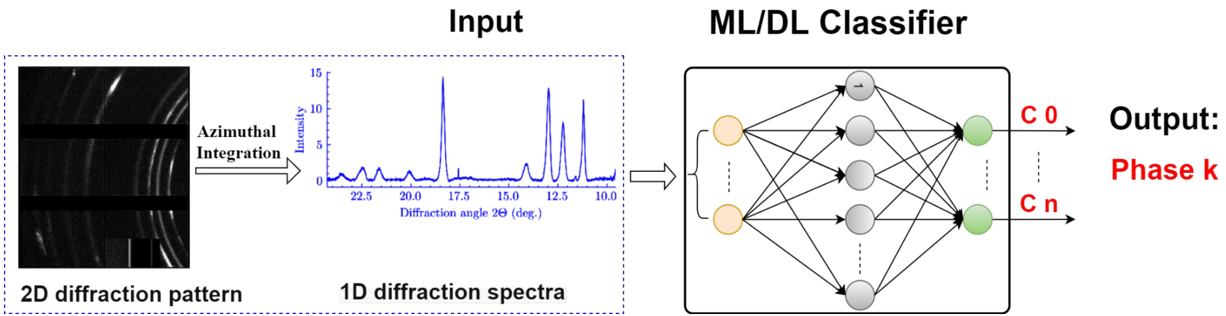


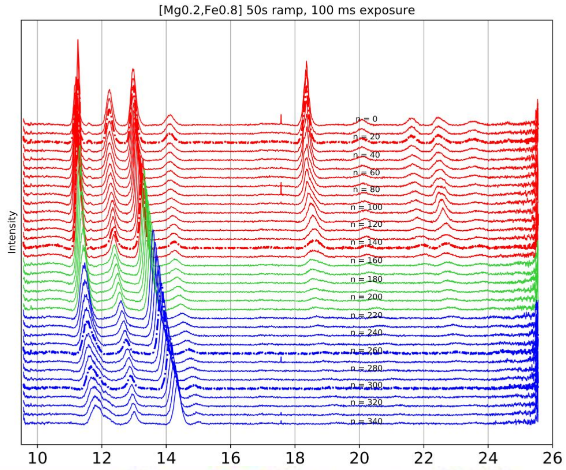
Condor Ref: DOI: 10.1107/S1600576716009213.

Machine Learning Applications

Yue Sun, Christian Plückthun, …Péter Hegedűs, Sandor Brockhauser
Data Analysis Group, University of Szeged

End-to-End ML Methods for Spectra Classification



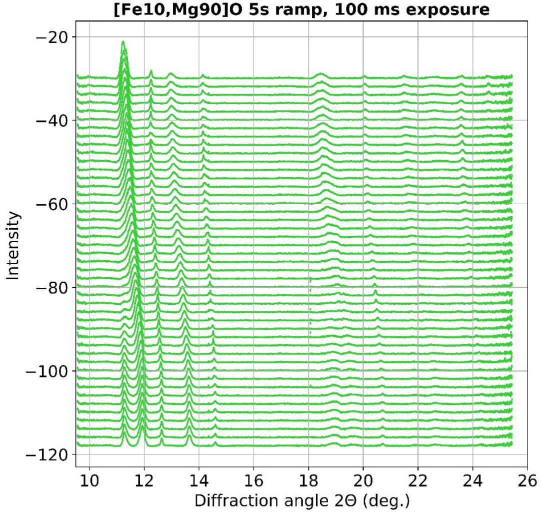


HED Diffraction Spectra data

Data set:
349 samples with each of 4023 features

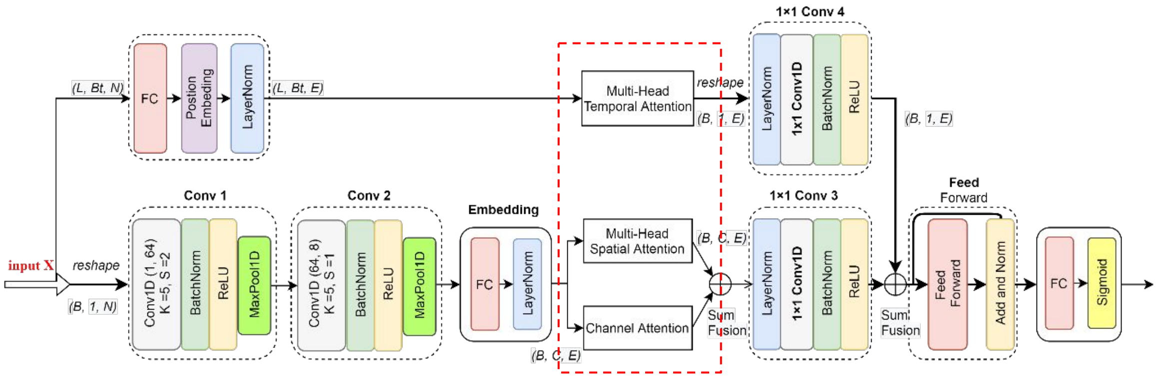
Two or Three phases:
Low and High pressure,
Phase transition

Other Spectra dataset



European XFEL

Convolutional SCT Attention network



Spatial Attention: $A_S = x + MultiHead(Q_S, K_S, V_S)$

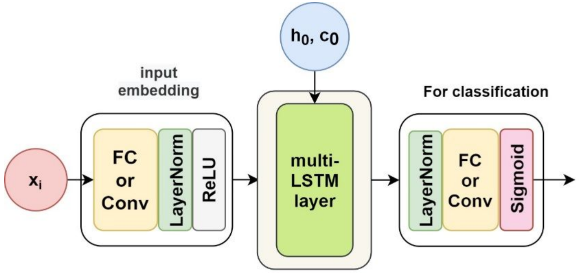
Channel Attention: $A_C = x + \alpha \cdot S_C \cdot V_c = x + \alpha \cdot softmax(Q_c K_c{}^T) V_c$

Temporal Attention: $A_T = x + MultiHead(Q_T, K_T, V_T)$

Sun, Y., Brockhauser, S. and Hegedűs, P., 2021. Comparing End-to-End Machine Learning Methods for Spectra Classification. Applied Sciences, 11(23), p.11520.

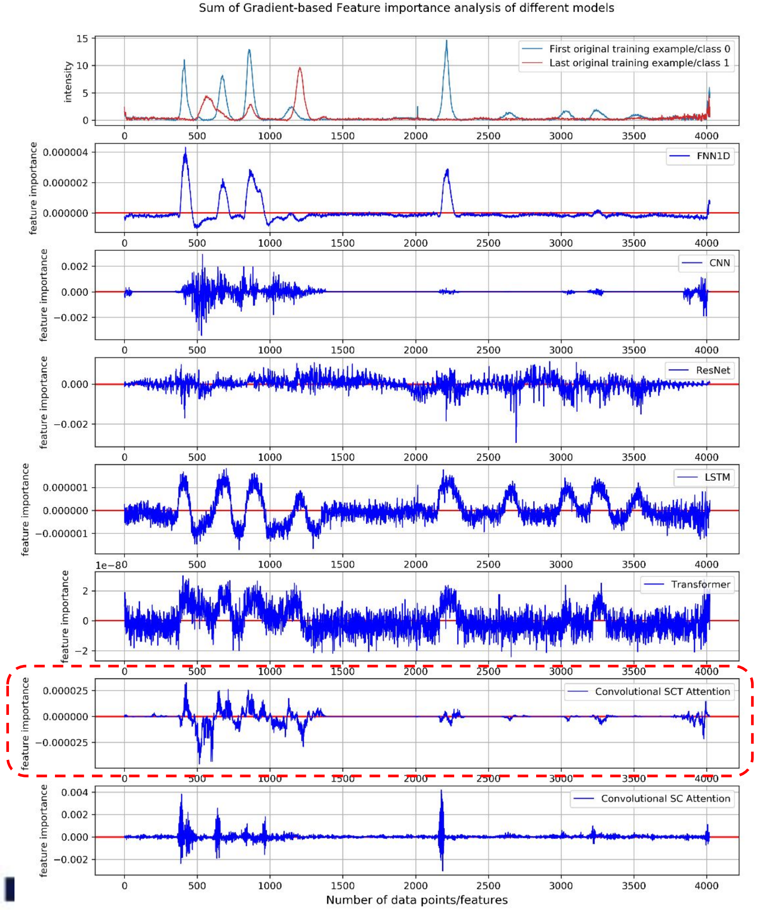
Other State-of-the-Art Deep Learning Approaches

- 1D FCNN Model
- CNN solution
- ResNets
- LSTM-based solution
- Transformer-based solution

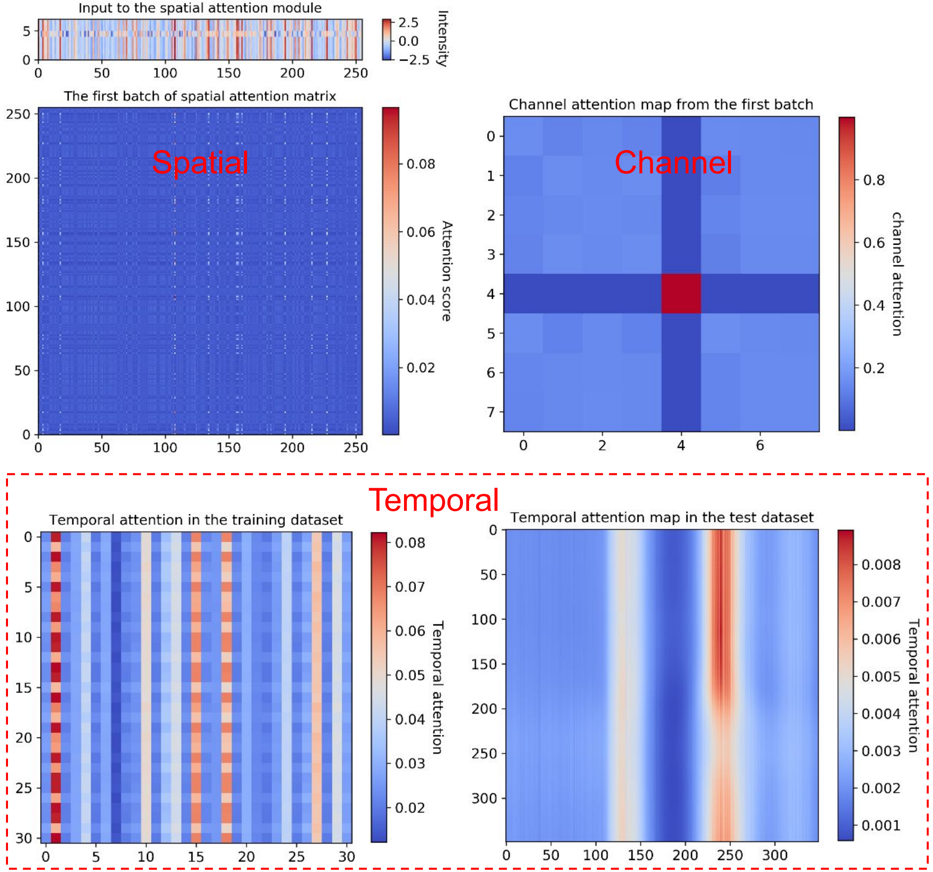


Multi-LSTM-based model

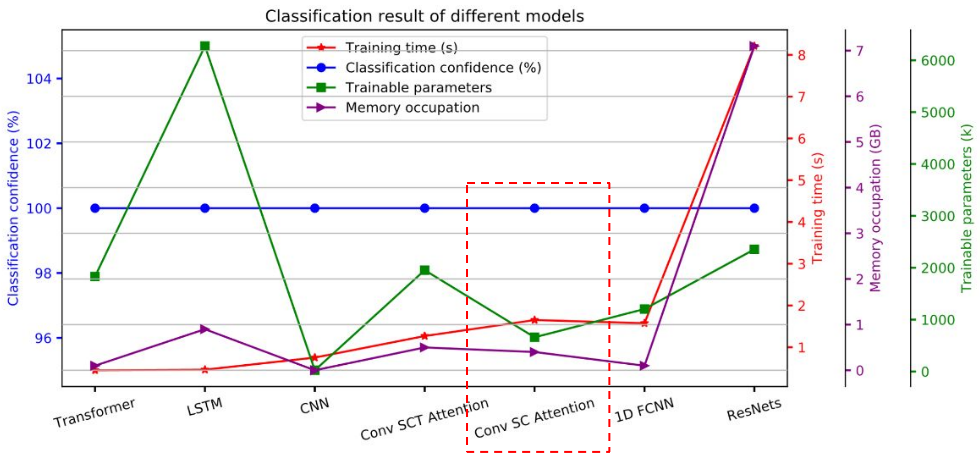
Feature Importance analysis based on Gradient Backpropagation



Qualitative Analysis of Self-Attention Scores in Convolutional SCT Attention model



Quantitative Model Evaluation



Model	Cls. confidence	Trainable Parameters	Training Time (s)	Memory (GB)	Epochs	ClS. boundary
1D FCNN	100%	1204335	1.576	0.1	300	(188,189)
CNN	100%	23307	0.753	0.0	5	(173,174)
ResNets	100%	2354535	8.208	7.1	6	(178,179)
LSTM	100%	6273391	0.463	0.9	25	(175,176)
Transformer	100%	1830511	0.449	0.1	30	(185,186)
Convolutional SCT Attention	100%	1953353	1.269	0.5	5	(188,189)
Convolutional SC Attention	100%	659525	1.652	0.4	8	(182,183)

Training samples: 1581 (31+50*31)

$$P_{conf} = 1 - \frac{C^1_{lower} - C^0_{upper} - 1}{N_t}$$

- ❖ All models can achieve 100% classification confidence;
- ❖ Transformer-based (0.449 s) methods consume the least training time;
- ❖ Conv SCT Attention model can suppress indistinguishable features better than others.

Summary

We are working on !

- ❖ Automation
- ❖ Image quality boosting
- ❖ Cluster your Big Data
- ❖ End-to-end ML methods for spectra classification

Potential future projects

- ❖ Object detection , Predictive Maintenance

Become our next pilot user !

Contact us today da@xfel.eu