

MExchange: bringing machine learning to scientific discovery

Presenter: Zhuowen Zhao

Advanced Light Source, Lawrence Berkeley National Laboratory

SMP Materials Seminar, 1/17/2023



Acknowledgement

MLExchange architecture team

Adam Green (LBNL, now at Apple)

Alexander Hexemer (LBNL)

Dylan McReynolds (LBNL)

Elizabeth Holman (LBNL, now at Stanford)

Hari Krishnan (LBNL)

Howard Yanxon (ANL)

Tanny Chavez (LBNL)

Tibbers Hao (LBNL)

Zhuowen Zhao (LBNL)

Collaborators

AIMM team

Apurva Mehta (SLAC)

Dani Ushizima (LBNL, CAMERA)

Ed Bernard (LBNL, Molecular Foundry)

Eric J. Roberts (LBNL, CAMERA)

Nicholas Schwarz (ANL)

Petrus H. Zwart (LBNL, CAMERA)

Ronald J. Pandolfi (LBNL, CAMERA)

Sergei V. Kalinin (ORNL)

Stuart I. Campbell (NSLS-II)

Subramanian Sankaranarayanan (ANL)

Funding

financial support by the U.S. Department of Energy through Collaborative Machine Learning Platform for Scientific Discovery (award No. 107514).

Table of contents

- Introduction (challenges)
- MLExchange architecture
- Live demonstration
- Use cases



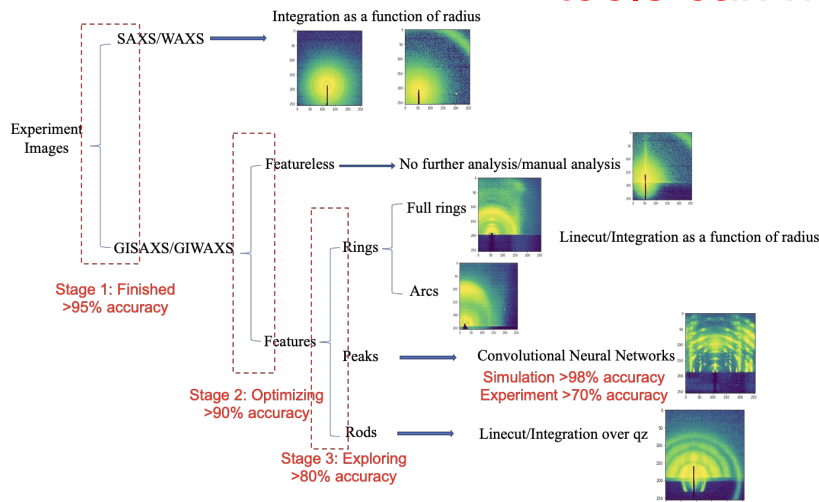
Challenges



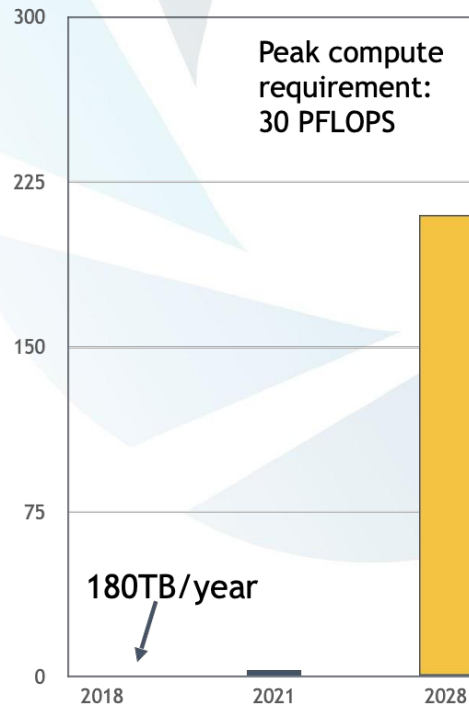
Data challenges for beamline users

- Big amount of data, growing annually
- On-the-fly analysis dealing with various data types (spectra, image data, etc.)
- Interactive visualization tools

Machine learning tools can help!



ALS Data Volume (PB/ year)



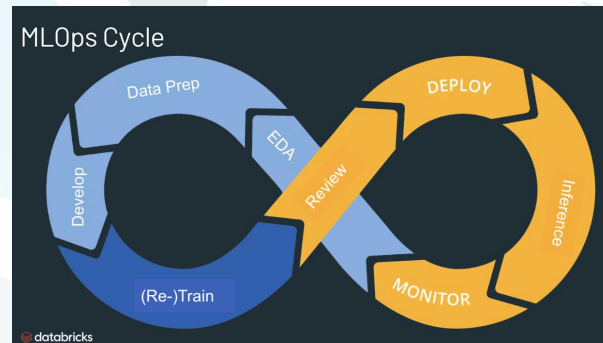
Challenges to use machine learning (ML)

- How to deploy and share ML workflows across DOE beamline facilities?



Challenges to use machine learning (ML)

- Many available ML tools require programmatic skills.
- Existing MLOps software:
 - not open-source (Weights & Biases)
 - do not have user system (MLflow)
 - do not have interactive interfaces customized for beamline users



General ML tool

pyMSDtorch

<https://pymstdtorch.readthedocs.io/>



Deep learning tools



Data processing tools



seaborn



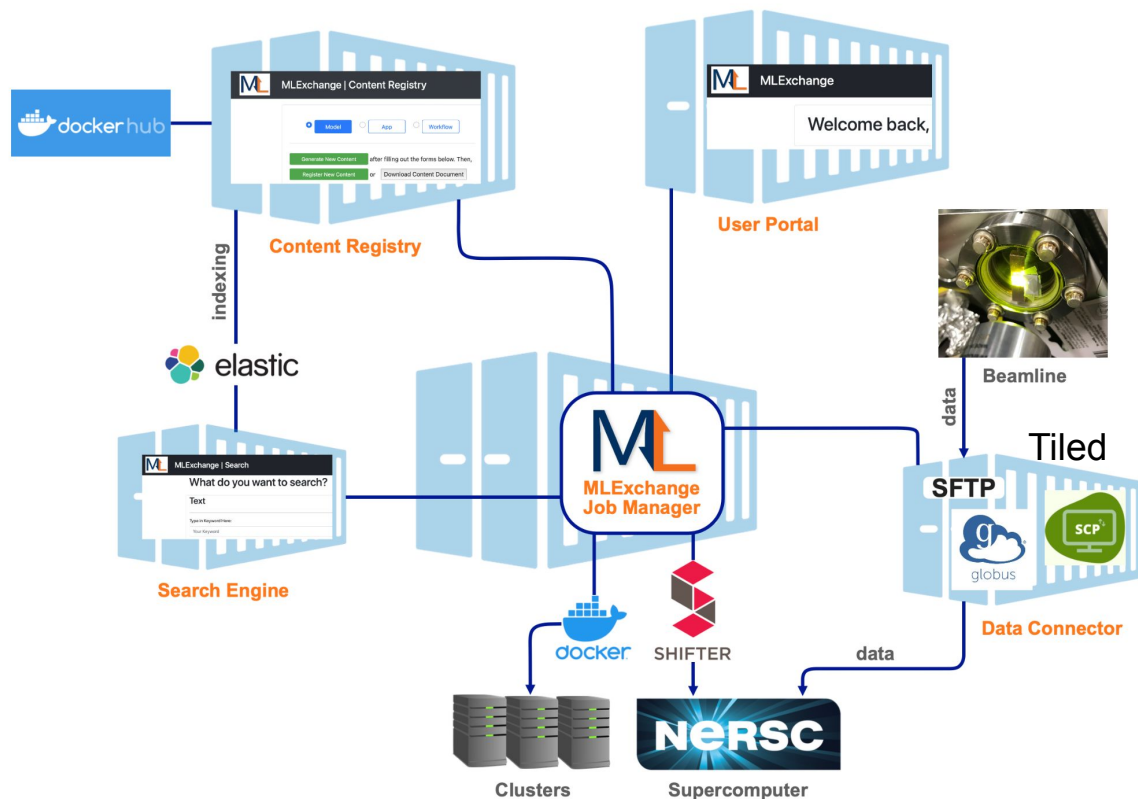
matplotlib

Visualization tools

MLExchange platform

MExchange architecture

- 5 major “containerized” components
- Communication through application programming interfaces (APIs)



Major components

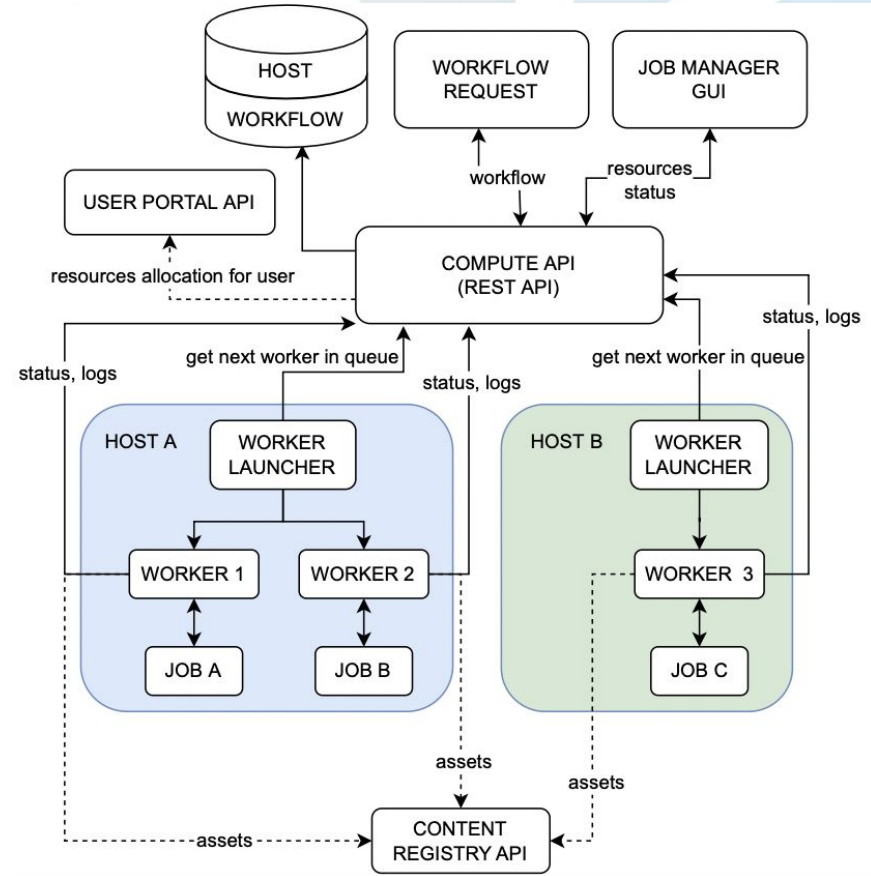


Job manager

- A central job coordinator: a queue based on the timestamp and available computing resources.



Image source: [Katerina Limpitsouni](#)



Content registry

MLExchange | Content Registry

Model App Workflow

Generate New Content after filling out the forms below. Then, Register New Content or Download Content Document

Please upload your content document after validation.
Validate Content Document Upload Content Document

Drag and Drop or Select Files

Please give a name for the model.
Enter model name.

Please provide the version for the model.
Enter model version.

Please provide the URI for the model.
Enter the URI.

Enter reference for the model.
Enter reference.

Enter the commands to deploy the model. Use comma to separate.
Enter commands to deploy the model

Resources requirement
CPU 0 GPU 0

+ Add GUI Component Double-click to See GUI Component(s)
Choose component type:
Select...

Refresh Model List	Delete the Selected	Launch the Selected
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pyC3B3	1.0.0	mlexchange team supervised mlexchange/pyc3b3
pyRNDtorch	1.0.0	mlexchange team supervised mlexchange/sdnetwork-notebook
pytorch_astorencoder	0.0.2	mlexchange team unsupervised mlexchange/unsupervised-classifier
random_forest	1.0.0	mlexchange team supervised mlexchange/random-forest-dc
TF-NeuralNetworks	1.0.1	mlexchange team supervised mlexchange/tfcoach

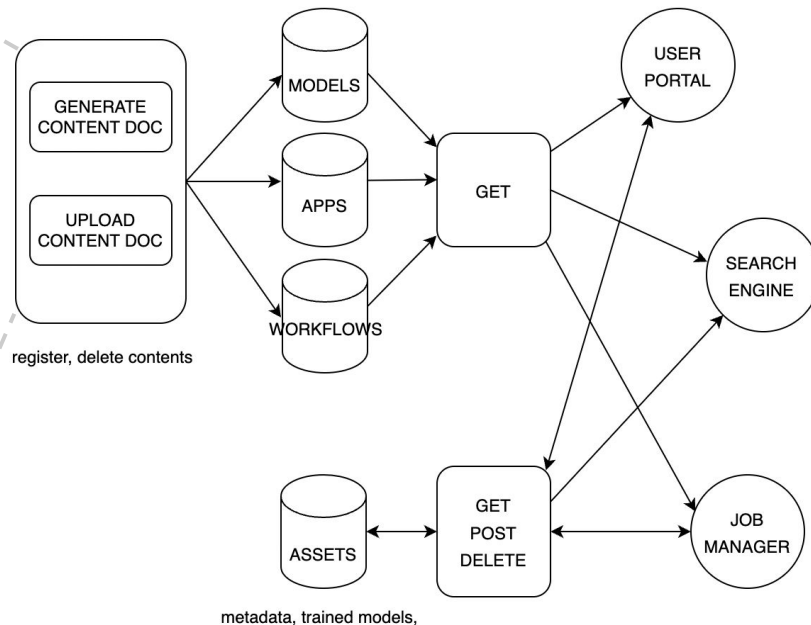
Refresh Job List	Terminate the Selected	Open the Selected Frontend App(s)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
description	service_type	submission_time
mlexchange/random-forest-dc	backend	2022-06-04T00:01:26.386000
mlexchange/random-forest-dc	backend	2022-06-04T00:01:26.386000
mlexchange/sdnetwork-notebook	backend	2022-06-04T00:01:26.386000
mlexchange/sdnetwork-notebook	backend	2022-06-04T00:01:26.386000

Content Registry GUI

Database

API

SERVICE



Content registry: adaptive GUI components

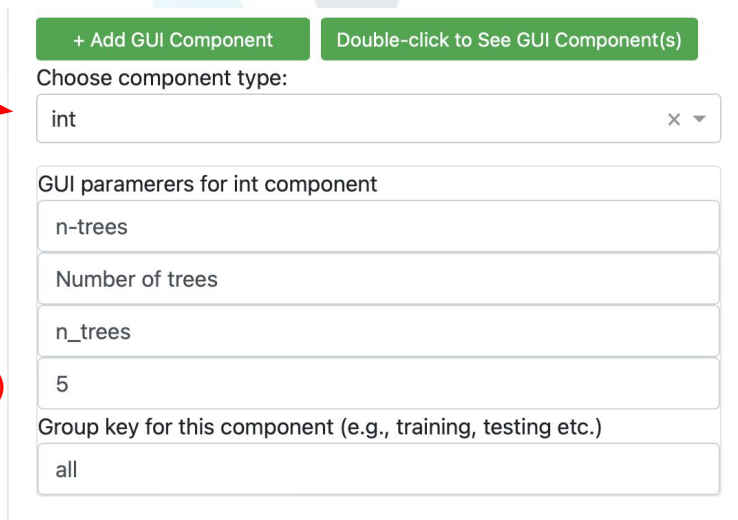
9 Dash component types

- [3 input forms](#) (int float, str)
- [Slider](#)
- [Dropdown window](#)
- [Radio items](#)
- [Boolean toggle switch](#)
- [Image](#)
- [Graph uploader](#)

component id
title (optional)

default value (optional)

group key (optional)



The screenshot shows a web interface for adding a GUI component. At the top, there are two green buttons: "+ Add GUI Component" and "Double-click to See GUI Component(s)". Below these is a dropdown menu labeled "Choose component type:" with "int" selected. Underneath, there are four input fields for "GUI parameters for int component": "n-trees" (with "Number of trees" as a label), "n_trees" (with "5" as a value), and "Group key for this component (e.g., training, testing etc.)" (with "all" as a value).



Automatically updates
GUI components in
frontend apps



The screenshot shows the rendered GUI component: a "Number of trees" label above a numeric input field containing the value "5".

User portal



- Create “user-role-group” relationship in a graph database (neo4j)
- Insert the user id and its authority into the access token after authentication (Google)
- Launch/terminate/go to apps on behalf of individual user



Welcome back, Zhuowen Zhao!

Launch Apps

Select apps to launch

Apps running status

Available apps

Launch

Select Applications

name	version	owner	uri	description
<input type="checkbox"/> color wheel	1.0.0	mlexchange team	mlexchange/colorwheel-notebook	xxx

Apps running status

Refresh List

Terminate the Selected

Open the Selected Frontend App(s)

description	service_type	submission_time	execution_time	job_status
<input type="checkbox"/> mlexchange/colorwheel-notebook	frontend	2022-10-18T18:45:47.779000	2022-11-02T23:05:46.720000	failed
<input type="checkbox"/> mlexchange/colorwheel-notebook	frontend	2022-10-18T18:45:47.779000	2022-11-02T23:10:45.306000	failed
<input type="checkbox"/> mlexchange/colorwheel-notebook	frontend	2022-10-18T18:45:47.779000	2022-11-02T23:41:03.842000	terminated
<input type="checkbox"/> mlexchange/colorwheel-notebook	frontend	2022-10-18T18:45:47.779000	2022-11-03T00:25:38.538000	terminated
<input type="checkbox"/> mlexchange/colorwheel-notebook	frontend	2022-10-18T18:45:47.779000	2022-11-03T16:53:08.276000	terminated

Team Memberships

This section exists to create project teams consisting of appropriate users who have registered and have been approved to use MLExchange resources. Please use the buttons below to navigate through team creation and team membership management. Note that the goal of teams is to serve as a method of controlling user access to owned assets relating to MLExchange.

Your Teams

Manage Members



Search

- Fast search for the contents stored in the content registry (based on user authority); it will evolve into a recommendation system.
- Search for similar images using pyCBIR¹ software

¹pyCBIR is Python-based tool for content-based image retrieval (Araujo et.al. 2018)

What do you want to search?

Text

Type in Keyword Here:

	content_type	name	version	type	owner	uri	application
<input type="checkbox"/>	model	PyCBIR	1.0.0	supervised	mlexchange team	mlexchange/pycbir	['image similarity search', 'cnn', 'recommen
<input type="checkbox"/>	model	Tensorflow-NN	1.0.0	supervised	mlexchange team	xxx	['mlcoach']

Image

Select Category Select CNN Select Searching Method Select Result image Num.

Upload Image Here:

List of Jobs

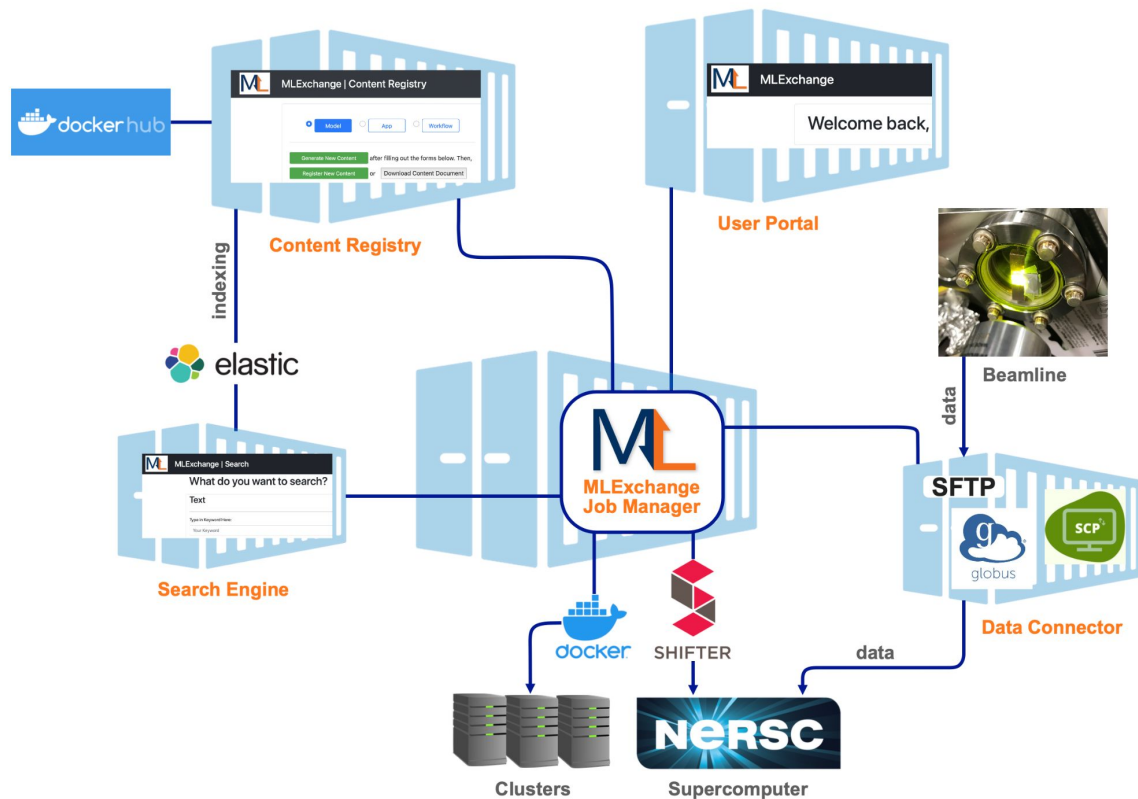
	Type	Status	Database	CNN	Searching Method	Number
<input checked="" type="radio"/>	deploy 20	complete	gisaxs	pretrained_vgg16	faiss	5
<input type="radio"/>	deploy 19	complete	gisaxs	pretrained_vgg16	faiss	7
<input type="radio"/>	deploy 18	complete	fibers	pretrained_vgg16	faiss	10
<input type="radio"/>	deploy 17	complete	fibers	pretrained_vgg16	faiss	10

Job Logs

```
2022-10-24 21:45:21.621299: E
tensorflow/stream_executor/cuda/cu
da_driver.cc:35] failed call to cuInit:
UNKNOWN ERROR (303)
pyCBIR started!
DBname: gisaxs
Model: TensorFlow_80
```

Revisit MLEXchange architecture

- Versatile deployment
- High modularity
- Scalability
- Easy accessibility



How-to live demo



How-to

Try MLExchange

GitHub: <https://github.com/mlexchange/mlex>

Docker Hub repository: <https://hub.docker.com/u/mlexchange1>

The official MLExchange portal: <https://mlexchange.als.lbl.gov> (Google authentication)

The development (beta) portal: <https://mlexchangebeta.als.lbl.gov> (Google authentication)

* Please note that Gmail account must be registered with both firstname and last name; o.w., you will see it keeps redirecting you back to the login page. The mlexchangebeta runs on our VM. We only have limited resources that support running max 6 (app) containers at the same time (automatically terminate in 5 mins).



How-to

Useful resources

MExchange documentation: <https://docs.mlexchange.als.lbl.gov>

MExchange website: <https://sites.google.com/lbl.gov/software-solutions-for-als/mlexchange>



Use cases

Use cases

- Image segmentation: a “standalone” frontend app

MLExchange | Image segmentation

Drag and Drop Here to upload!

Then Choose Dataset
Bead Experimental

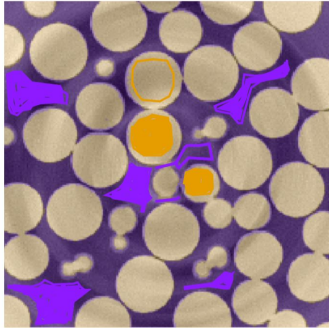


Image Slider
0 199

List of Jobs

Type	Status	Dataset	Model	Parameters
deploy 15	complete	data/bead_pack.tif	random_forest	training 17
training 17	complete	data/bead_pack.tif	random_forest	{ 'n_estimators
deploy 14	complete	data/bead_pack.tif	random_forest	training 16
training 16	complete	data/bead_pack.tif	random_forest	{ 'n_estimators

Annotation Tools

Label class
0 1 2 3 4

Width of annotation paintbrush
4

Delete All Annotations

Model

Choose Segmentation Model
Random Forest

Model Source
Adapted from Dash Plotly image segmentation example

Number of Trees
30

Tree Depth
8

Choose Deployment Location
Local (cpu)

TRAIN TEST

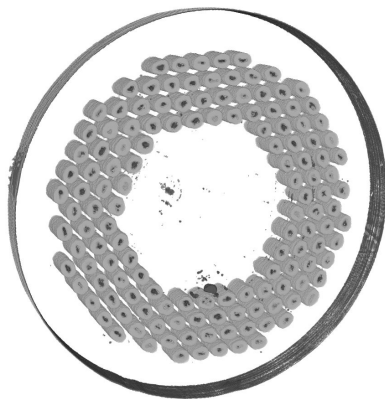
Show segmentation

Job Logs

```
classified 2
classified 0
classified 1
classified 3
classified 4
classified 5
classified 6
```

APS tomography data

- Segmented area visualization (ImageJ 3D viewer) by Jean-Francois Croteau



Uploaded: XX_APC_R.tif

Then Choose Dataset
XX_APC_R.tif

Annotation Tools

Label class

0 1 2 3 4

Width of annotation paintbrush

3

Delete All Annotations

Model

Choose Segmentation Model
Random Forest

Model Source
Adapted from Dash Plotly image segmentation example

Number of Trees
30

Tree Depth
8

Choose Deployment Location
Local (cpu)

TRAIN TEST

Show segmentation

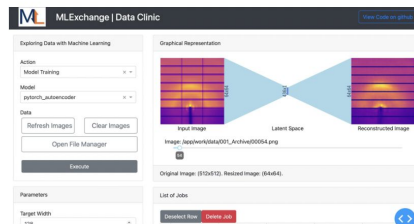
Image Slider
0 854

List of Jobs

Use cases

- Image labeling pipeline: a combination of multiple frontend apps

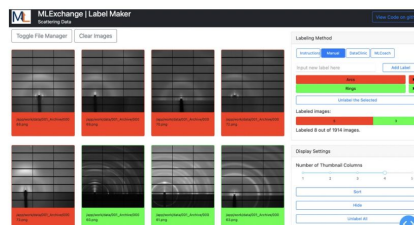
Data Clinic



self-supervised/unsupervised



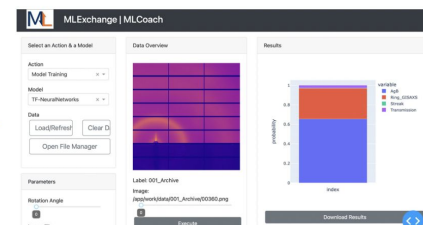
Label Maker



Labeled Data



MLCoach



supervised



Thank you !

Q & A

Contact: zzhao2@lbl.gov