Explainability in Satellite Based Remote Sensing of Nuclear Facilities

Miguel Avalos

Texas A&M

02/09/2023

Introduction and Motivation

- To provide a platform for passive or on-demand monitoring of nuclear facilities to evaluate local processes as well as have potential for nuclear security anywhere within the earth’s atmosphere.

- With the limited amount of data, the goal is to create AI that can accurately classify nuclear facilities and their processes, and making this AI reliable and trustworthy enough to be used in real situations for nuclear security purposes.

Mission Relevance

- This project reduces the threat of nuclear and radiological terrorism by surveying and automatically characterizing phenomena of interest

- Preventing nuclear weapons proliferation and reducing the threat of nuclear and radiological terrorism around the world are key U.S national security strategic objectives that require constant vigilance. NNSA's Office of Defense Nuclear Nonproliferation works globally to prevent state and non-state actors from developing nuclear weapons or acquiring weapons-usable nuclear or radiological materials, equipment, technology, and expertise.

  - *Website:* https://www.energy.gov/nnsa/missions/nonproliferation

Methodology

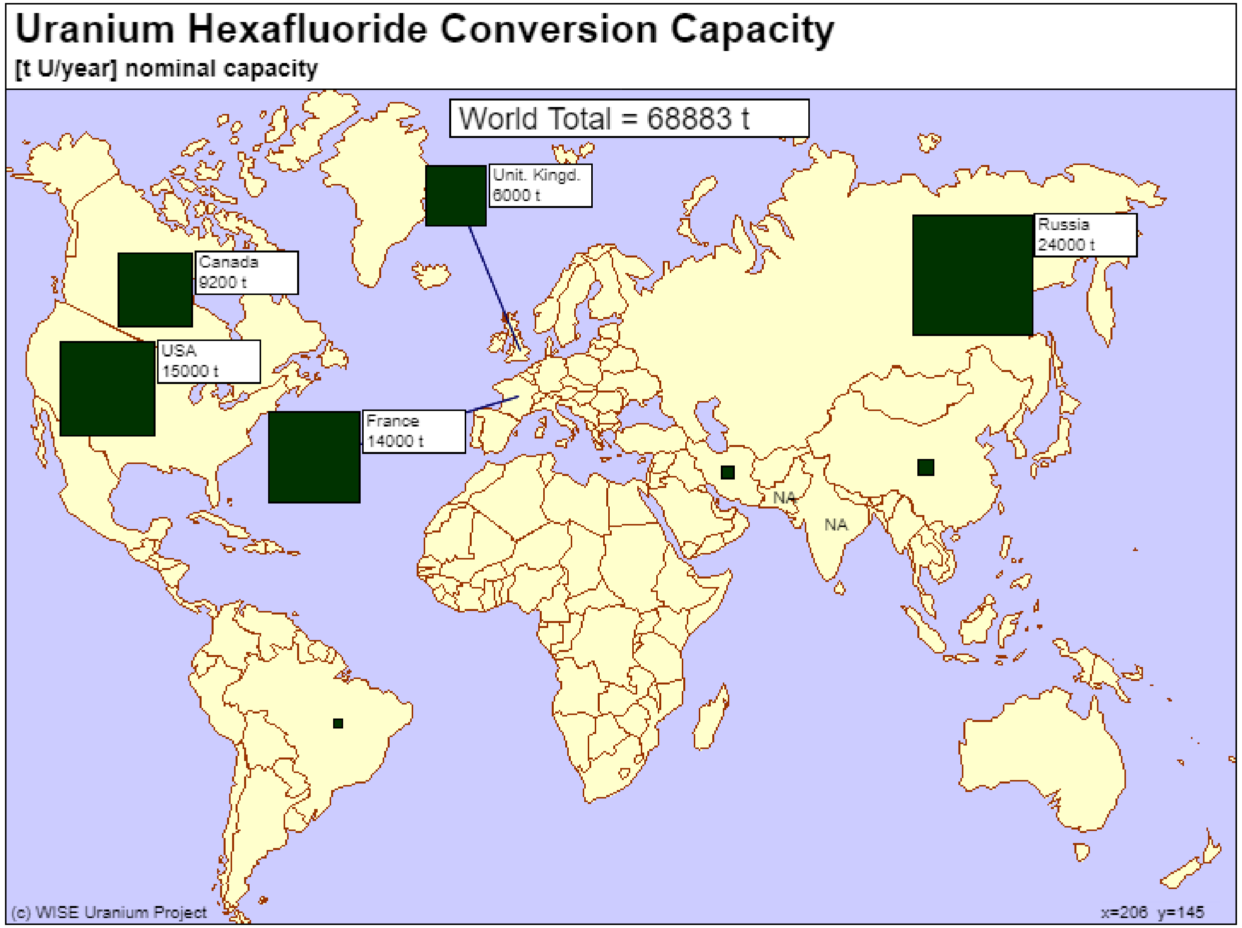
- Look into what state of the art deep learning and remote sensing can offer in security options for nuclear facilities. Such as detection of clandestine facilities, or monitoring current ones.

- Use what is openly available such as Google Earth and other satellite data. Some will be blurred or unavailable.

- Get as much image data from this as possible so that we can work with deep learning algorithms. Reactor, mining, conversion, enrichment, fuel fabrication, storage, and reprocessing.

Methodology

- Obtain as much data about the facilities build, location, and processes so that it can be used for understanding how these facilities work and where they are.
- World Information on Energy Uranium Project, Wikipedia, U.S. NRC, IAEA, World Nuclear Association. Great sources for information on nuclear facilities
- Use deep learning to classify, and use explainability to determine how well the methods were done, and how much trust can be put into this machine.



National Nuclear Security Administration

Technical Work and Results

- Machine Learning
  - Working with Satellite Images, using Deep Learning
  - Necessary for Mining, Conversion, Enrichment, Fuel Fabrication, Reactors, and more.
  - Data is limited by Conversion Facilities
  - Few-Shot algorithms
- Explainability
  - What is our machine learning
  - What do we expect it to learn
  - Is what it learned make sense

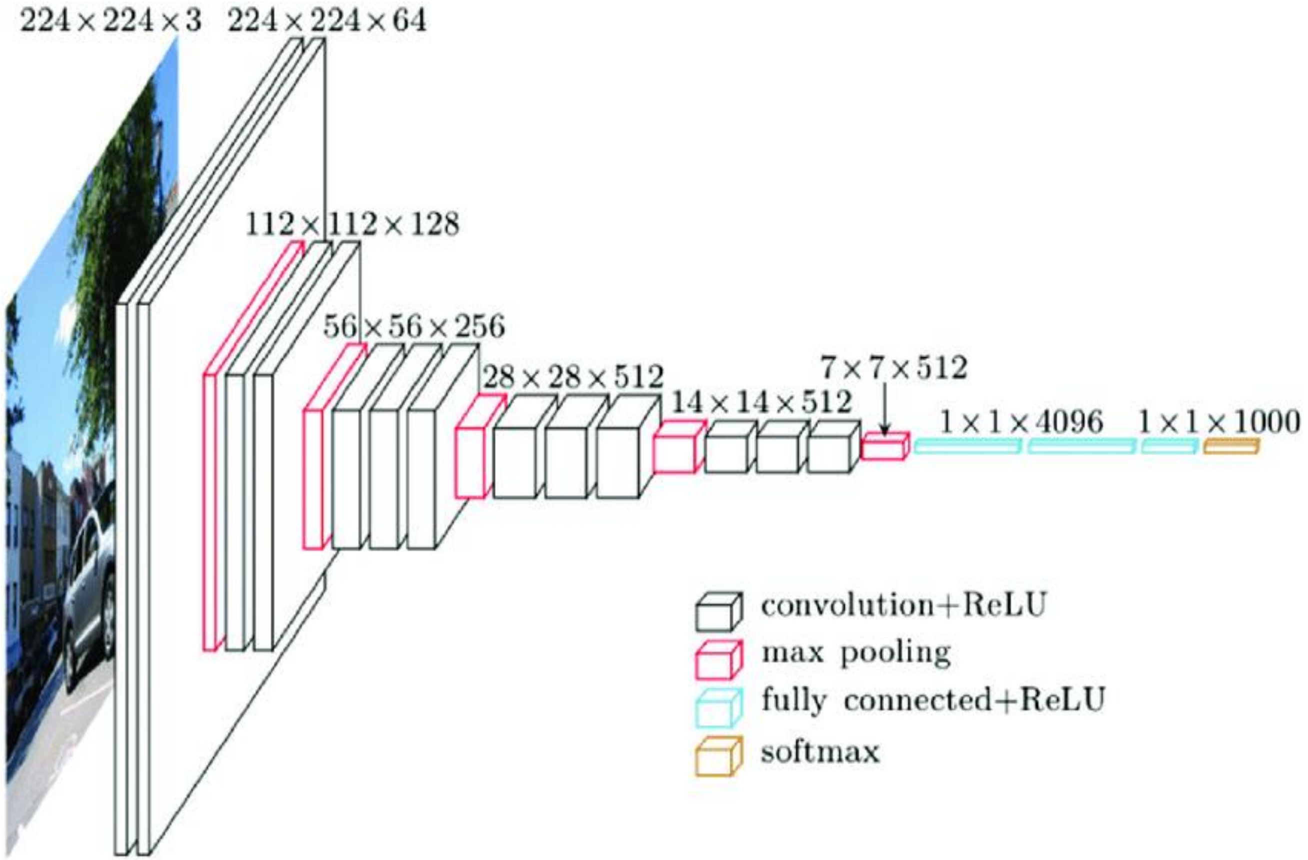
Deep Learning

Classic Deep Learning Convolutional Neural Networks(CNN)

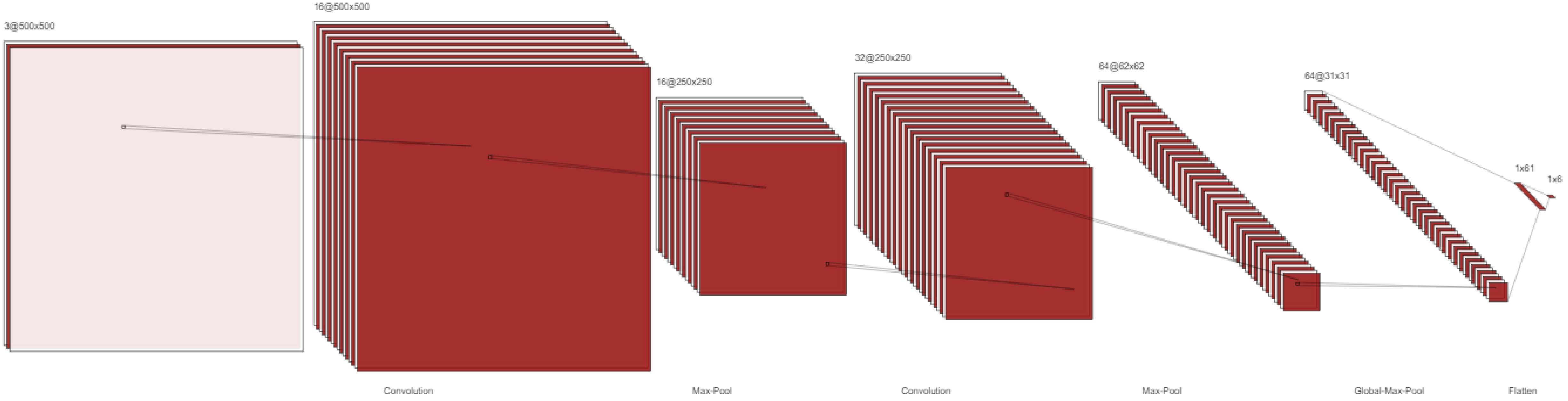
Requires lots of data to be effective(thousands)

Testing to see if it can recognize anything from this data

~40-50% accuracy(difficult to trust with less than 500 images total, and only 5 unique images per class)



Deep Learning



Limited Data

12 total Conversion facilities, with half being unavailable. Plenty of mining facilities, but with different methods, leading to similar issues

Requires different methods to solve this if we want to make a classifier

Create data, work with facilities or satellites to obtain data in different conditions

Port Hope, Ontario*
Pierrelatte*
Isfahan
Springfields, Lancashire
Metropolis, Illinois

Not all convert from U_3O_8 to UF_6

Limited Data

Data augmentation: Orientation, flipping, zooming, noise, occlusion

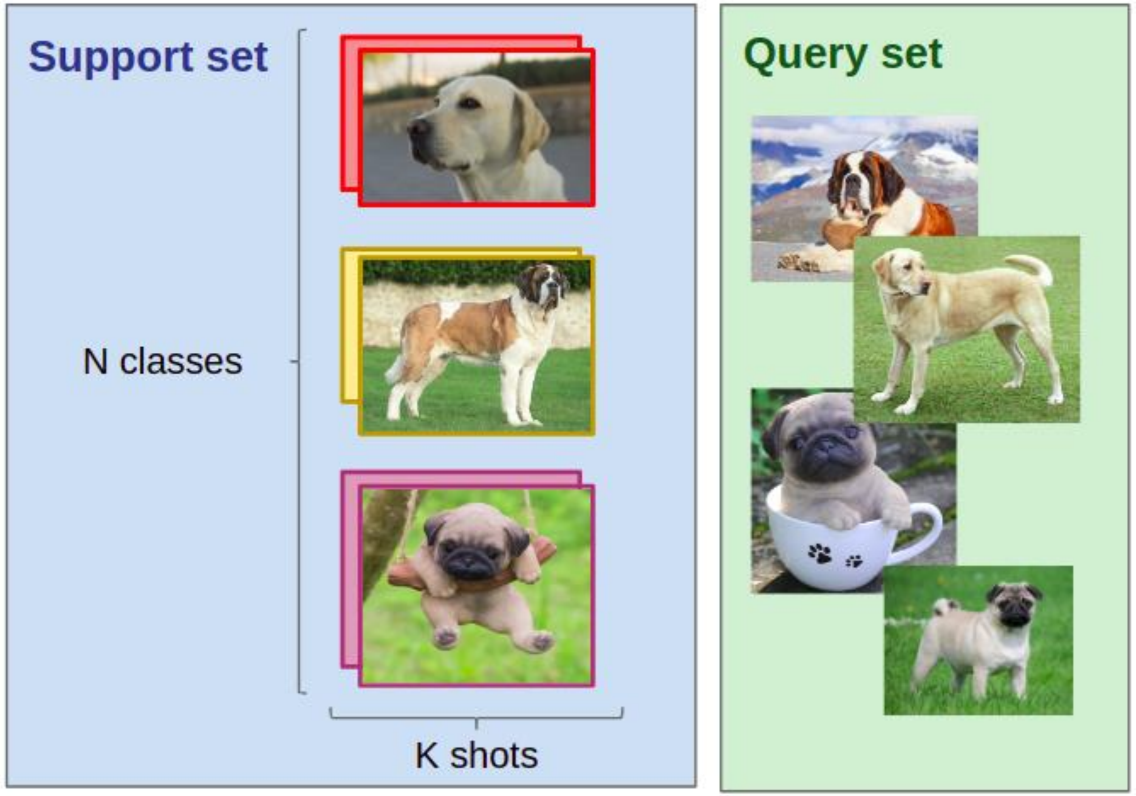
Get more information on what is necessary at these facilities

Few-Shot

Uses a support set of images to compare to the query set to obtain features

Requires much less Data (Less than 10/label)

Uses CNN’s to convert images into a features space, and classifies them through this method

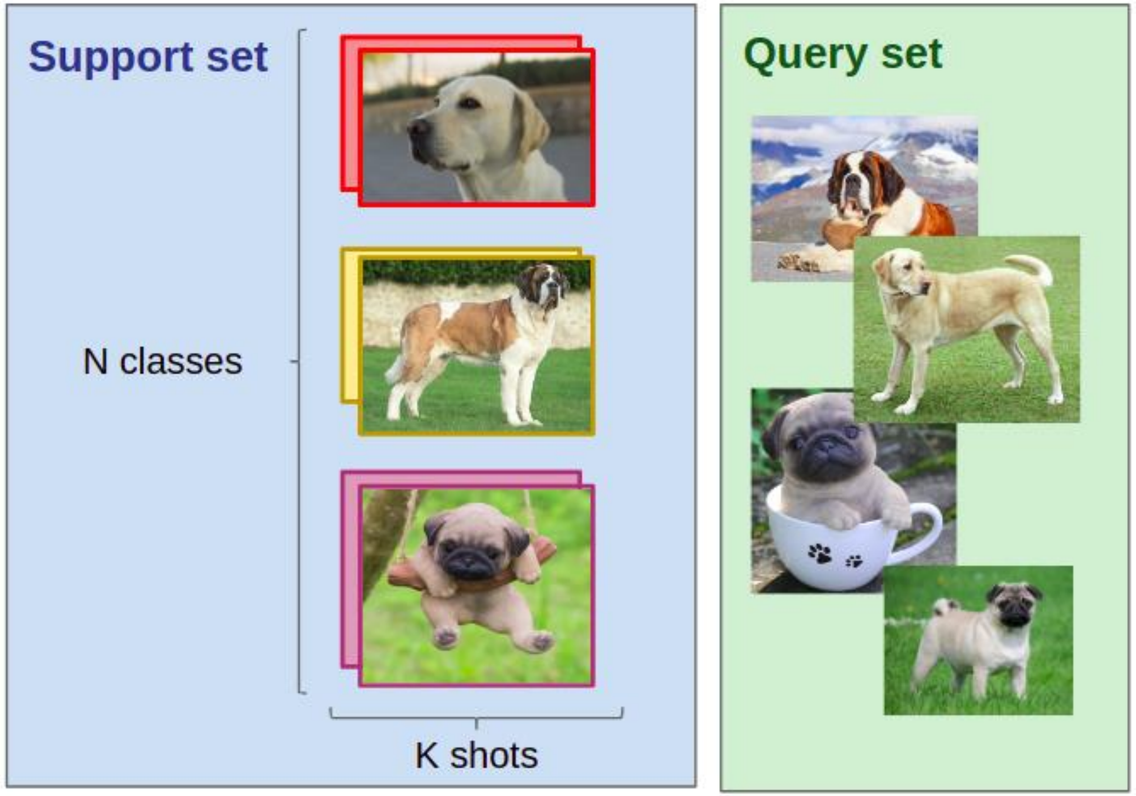


Few-Shot

Much better results for complicated tasks with less data

Simulates a type of memory that's not present in classical deep learning

Still needs a decent amount of testing data

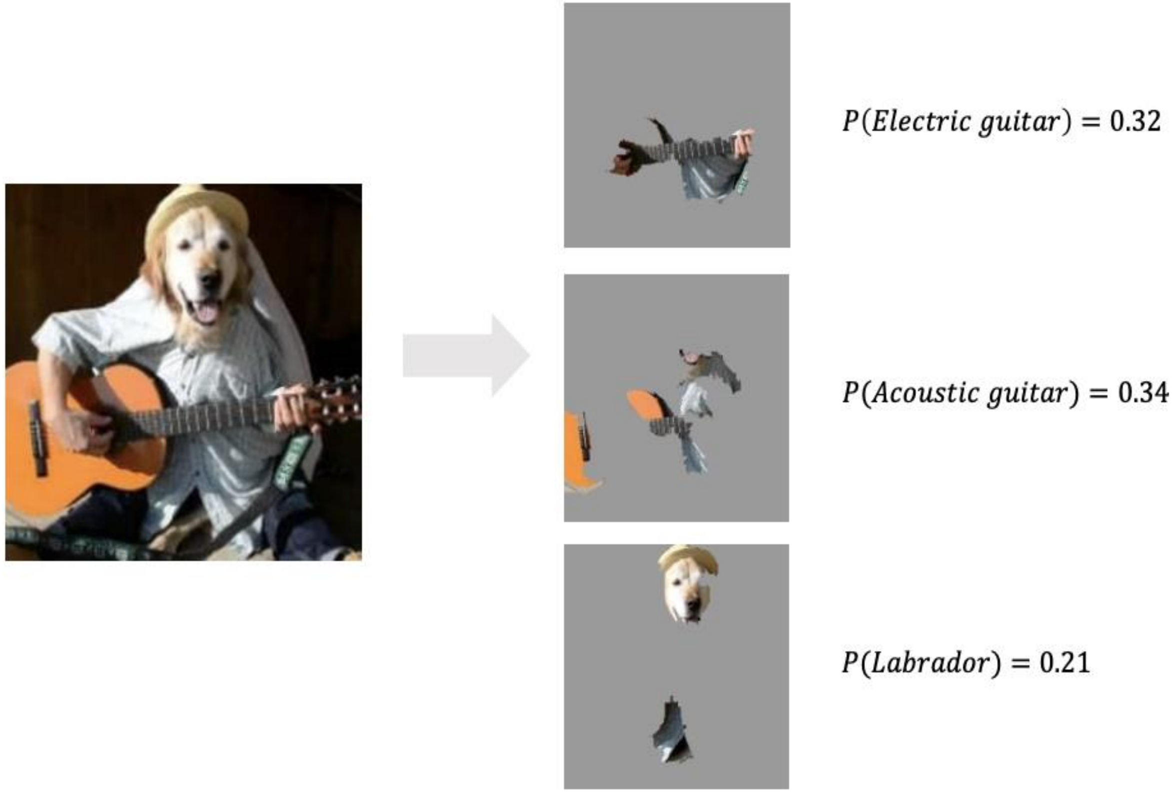


Explainability

What is a machine learning, and how can it be less of a black box for classification that requires explanation or confidence

Programs are built to look at images and explain why a certain classification was given

Gives user information about what the machine has learned and can make adjustments if needed, or explain with proof what the product does to a client



Explainabilit

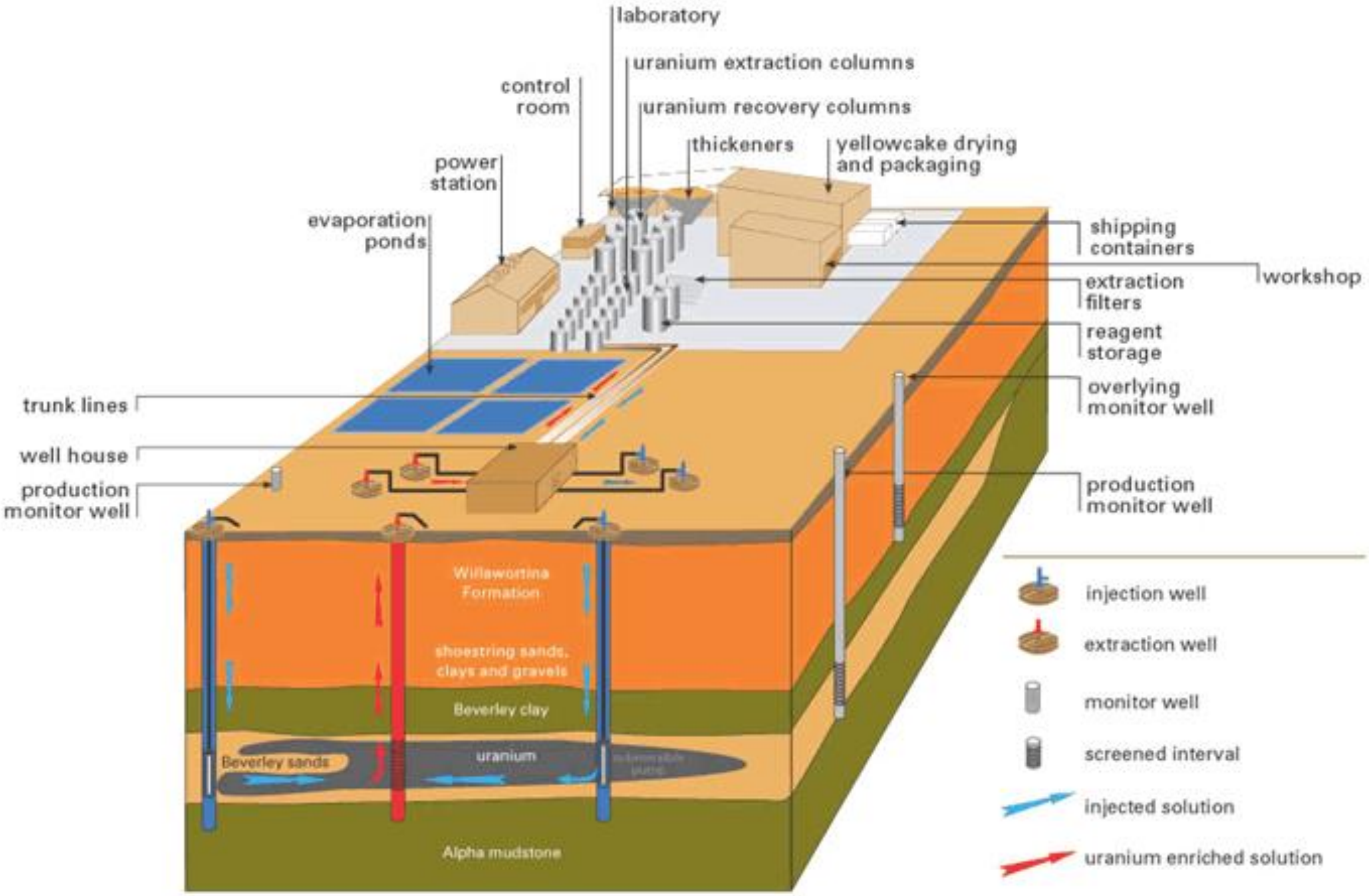
In-Situ Leaching mining Facility

Evaporation ponds

Trunk/Pipe lines

Well Field

Desert or sandy environment



National Nuclear Security Administration

Explainability

In-Situ Leaching mining Facility

Evaporation ponds

Trunk/Pipe lines

Well Field

Desert or sandy environment

Explainability

In-Situ Leaching mining Facility

Evaporation ponds

Trunk/Pipe lines

Well Field

Desert or sandy environment

National Nuclear Security Administration

Explainability

Conversion

What are the unique features?

Not all are U3O8 -> UF6

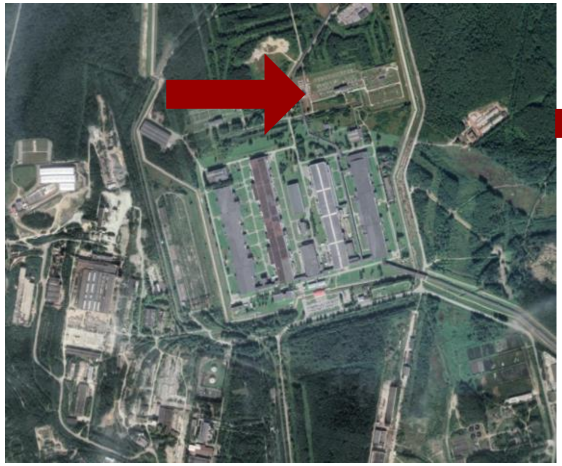
Explainability

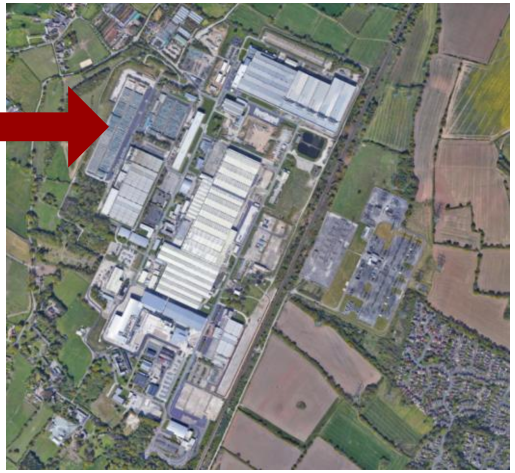
Enrichment

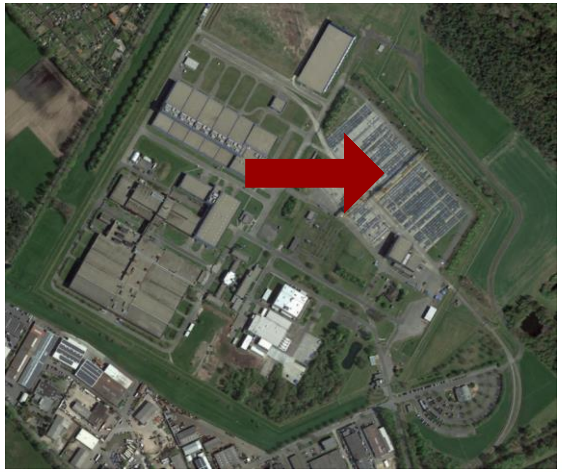
Lots of storage containers visible

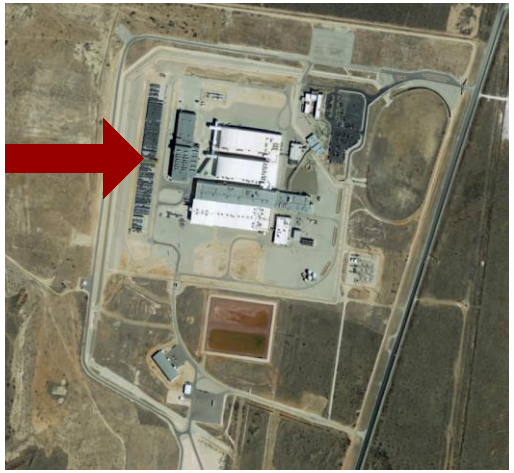
2 of them have pools

Buildings of same size and shape very close together









Explainability

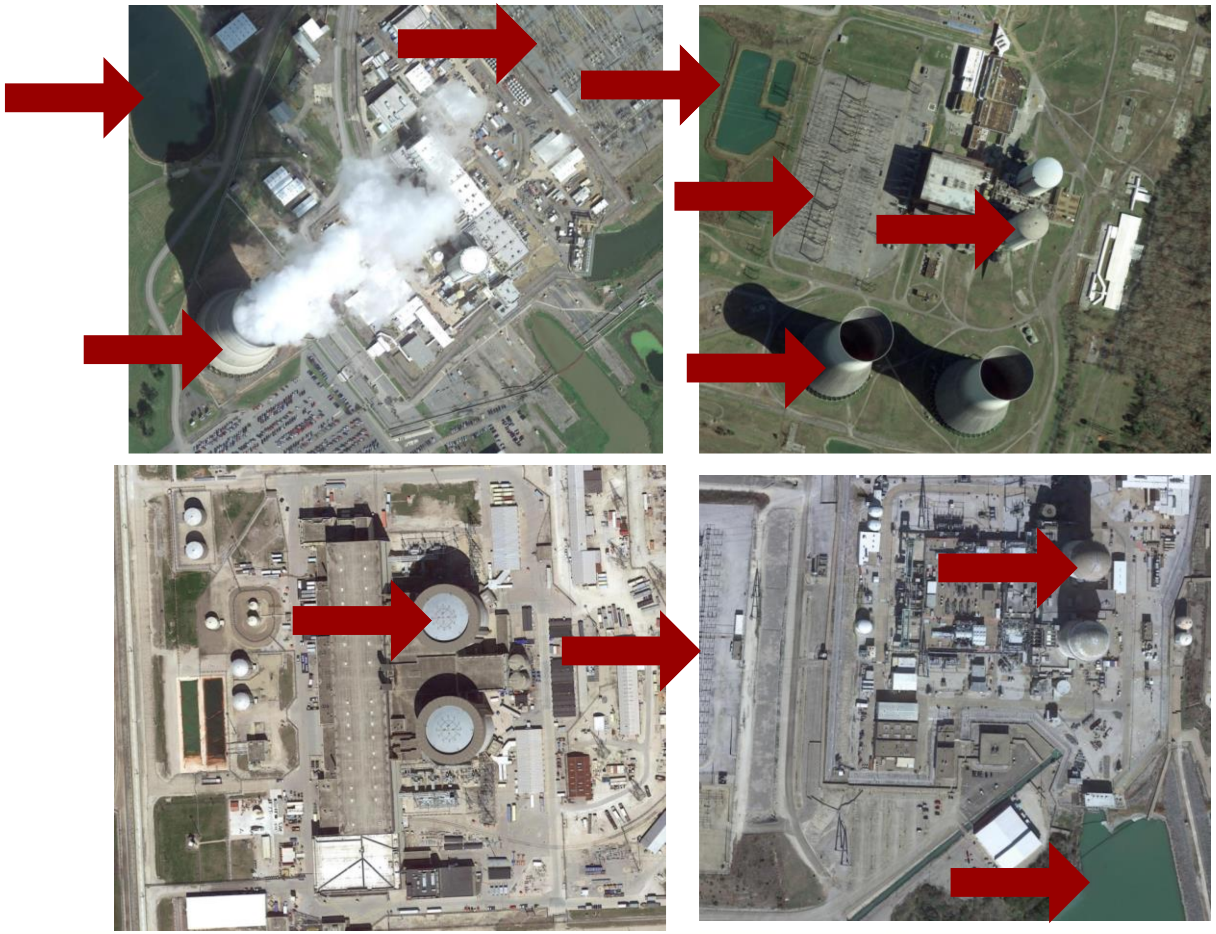
Reactor

Cooling Tower

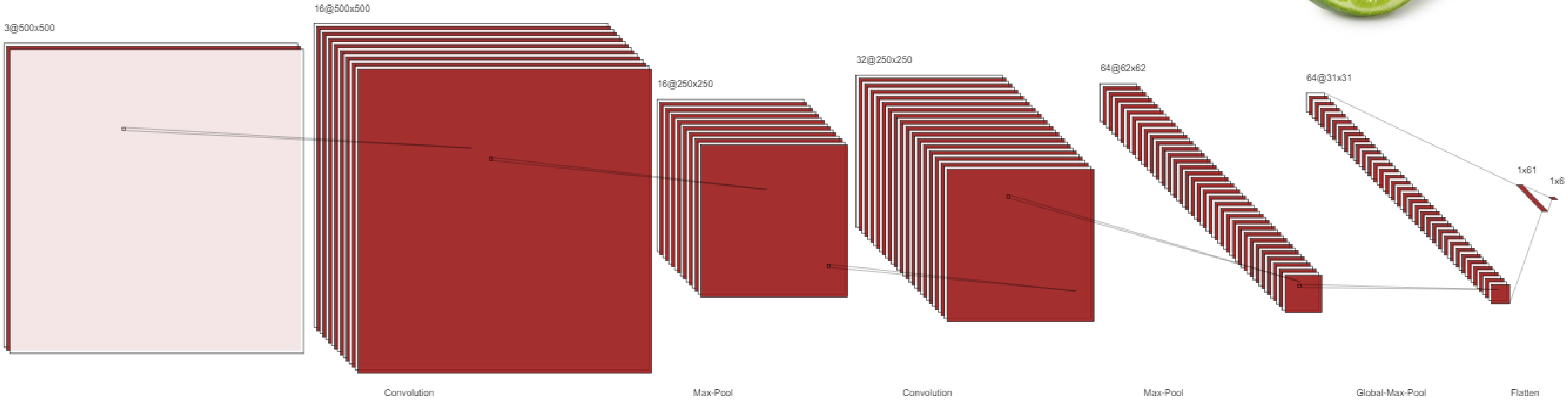
Switchyard

Large circular buildings

Near water



Explainability



ETI Impact

- Research alongside PNNL
- Workshops with undergraduates
- Networking with others in similar fields
- Connections with labs
- Personnel transitions: Continuing work with PNNL

Conclusion

- Classical deep learning shows that there are features to learn from, but is insufficient to classify due to a lack of data

- Most of the facilities have unique features that could make them distinct from other buildings or facilities, but there are limitations

- The data so far shows that it is difficult to use these methods to classify ALL nuclear facilities, and therefore may need more data or better methods for classification. Based on what is studies from Few-Shot methods, it seems like a promising approach for monitoring.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Department of Energy / National Nuclear Security Administration under Award Number(s) DE-NA0003921.