A Multimodal-Deep Learning System for the Retrieval of Open Source Data on Nuclear Proliferation Activities

Yana Feldman, Margaret Arno, Carmen Carrano,
Brenda Ng, and Barry Chen

LLNL-PRES-755031

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

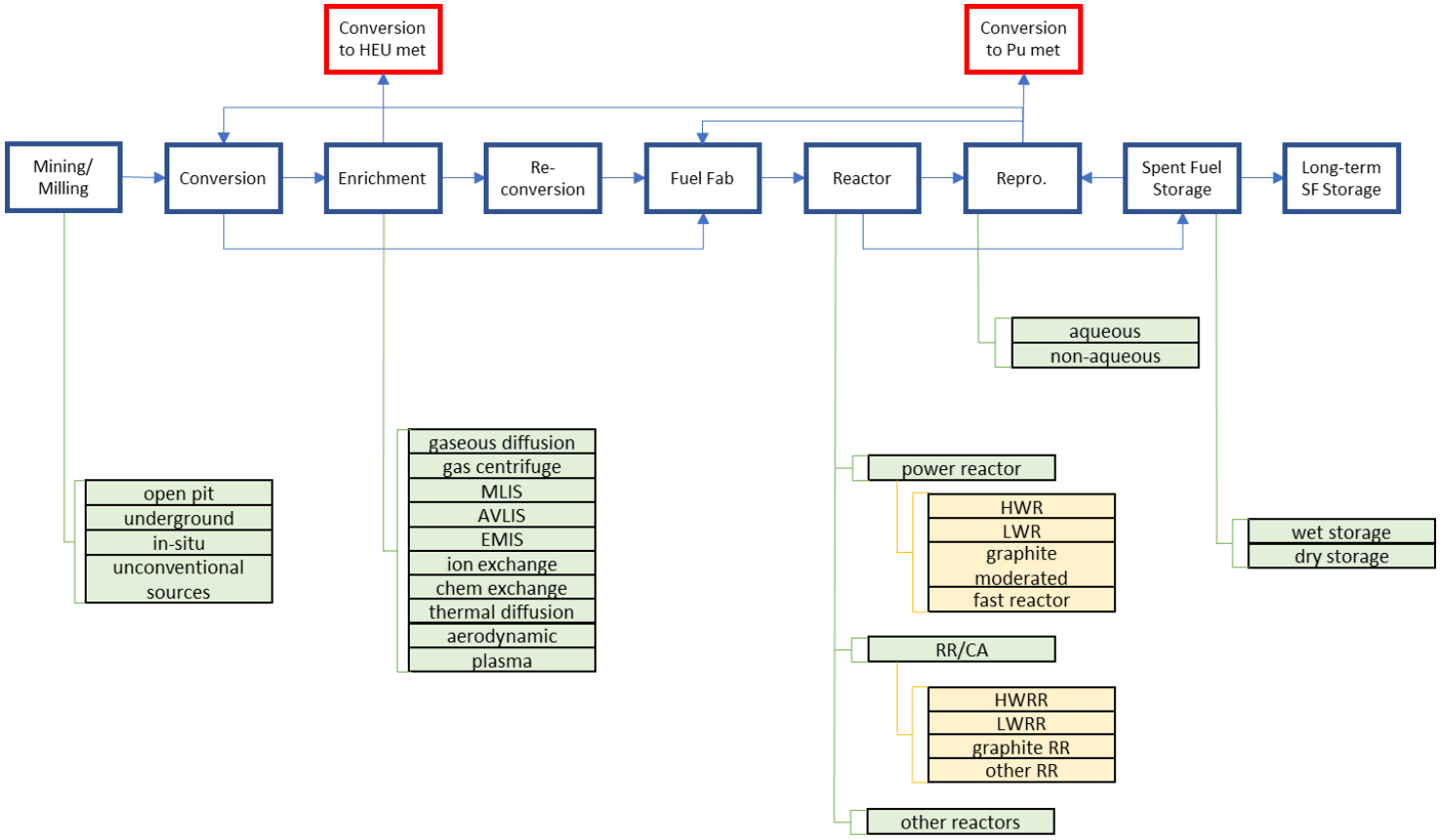
Agenda

- Analyst challenges
- Process model
- Developing a large scale multimodal information retrieval system
  - Datasets
  - Training
- Retrieval results
  - Image-to-image
  - Stock image-to-real image
  - Text-to-image
  - Image-to-video

Analyst Challenges

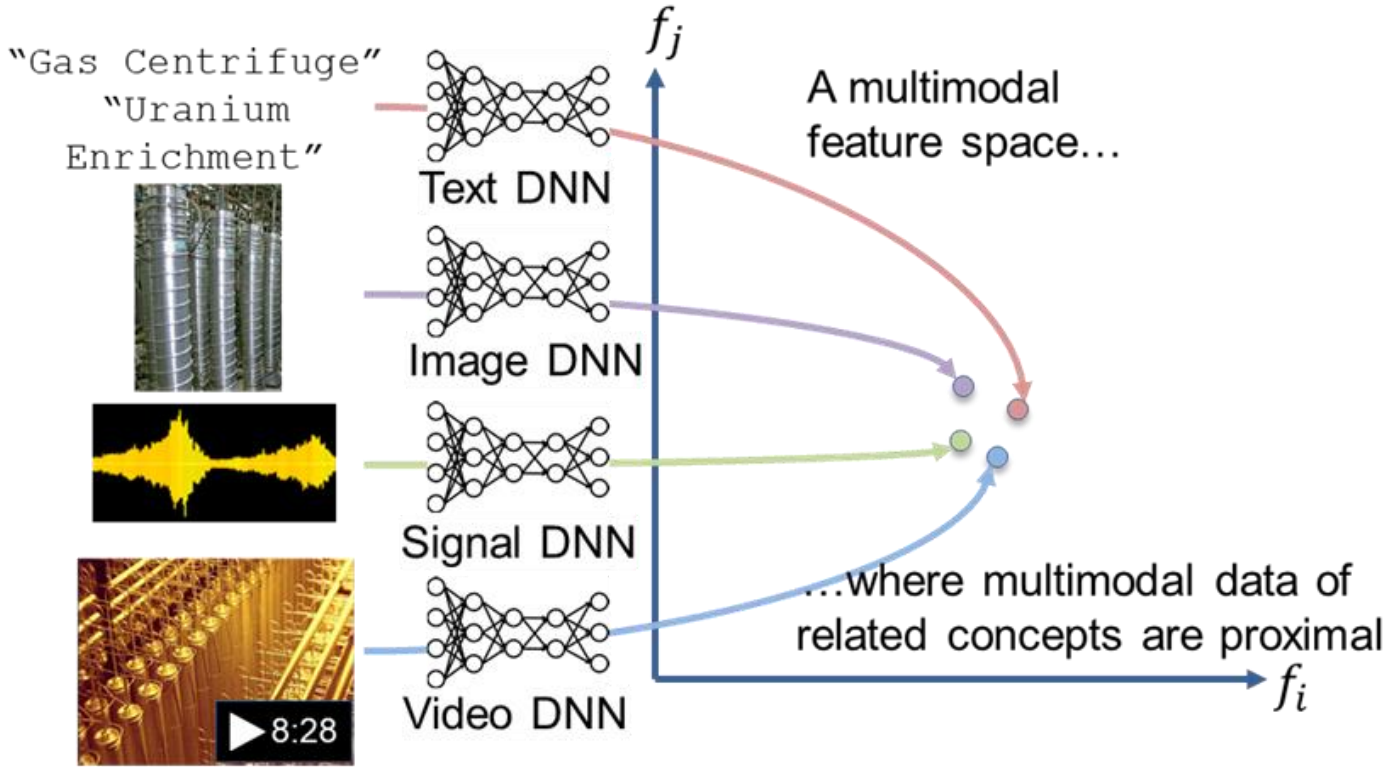
- Increasing volume, rate, and variety of information
- Potential for low density of highly valuable data
- Unlabeled or improperly labeled data
- Increasing number of sources that are difficult to monitor systematically
- Unstructured data

Process Model



Large-Scale Multimodal Information Retrieval System

Couples unsupervised and supervised pre-training on large-scale, generic open source datasets with fine-tuning on small, curated, nonproliferation-specific datasets

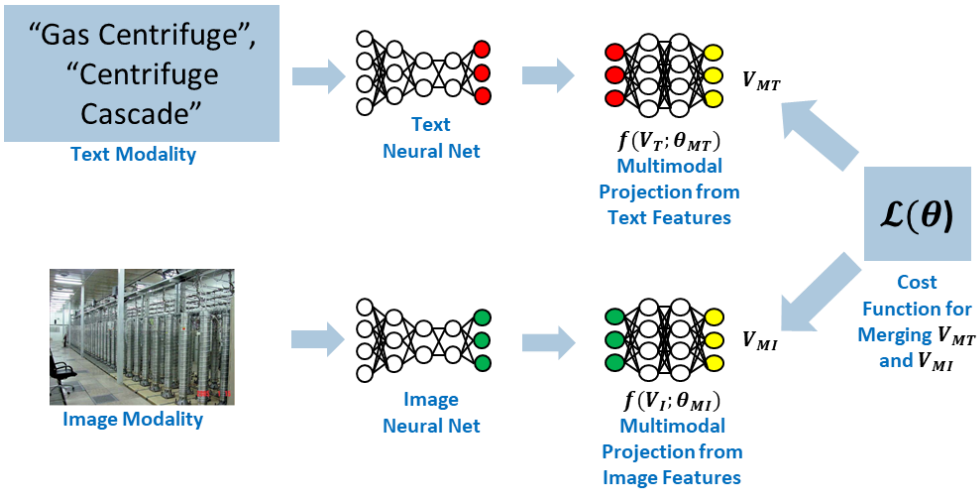


Step 1: Training Unimodal DNNs

Modality		Model	Training Data
Text	**“Gas Centrifuge”** **“Centrifuge cascade”**	FastText	New York Times snapshot (1987-2007) Wikipedia snapshot (April 3, 2015) Subject matter expert (SME)-curated NFC-related documents
Image		Imagenet-21k-inception	ImageNet SME-curated NFC images
Video		Treated as a series of static images	SME-curated NFC videos

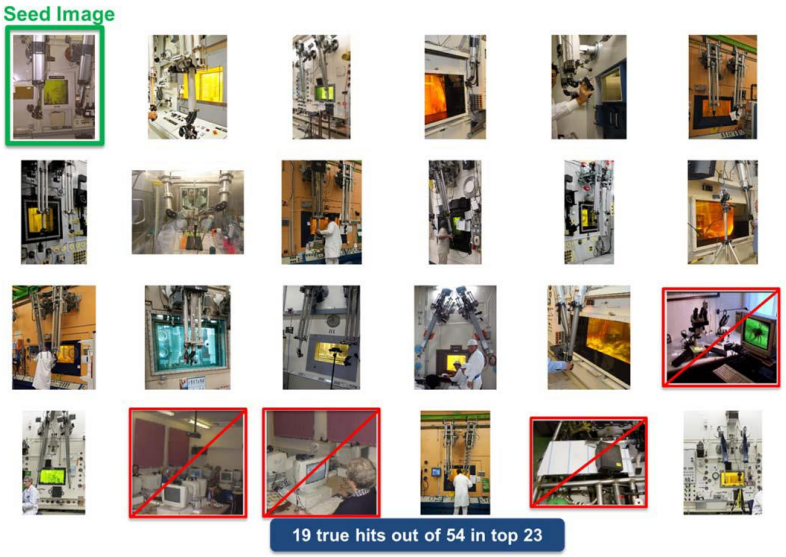
Step 2: "Merger" Mappings

- Training data: need aligned data from different modalities
  - training partition of ESP-Game dataset (53,252 images and corresponding tags)
  - 12 labeled examples from each object category in NFC-related image datasets
- Enables retrieval of items by returning nearest neighbors to a seed query in the multimodal feature space



Retrieval Results

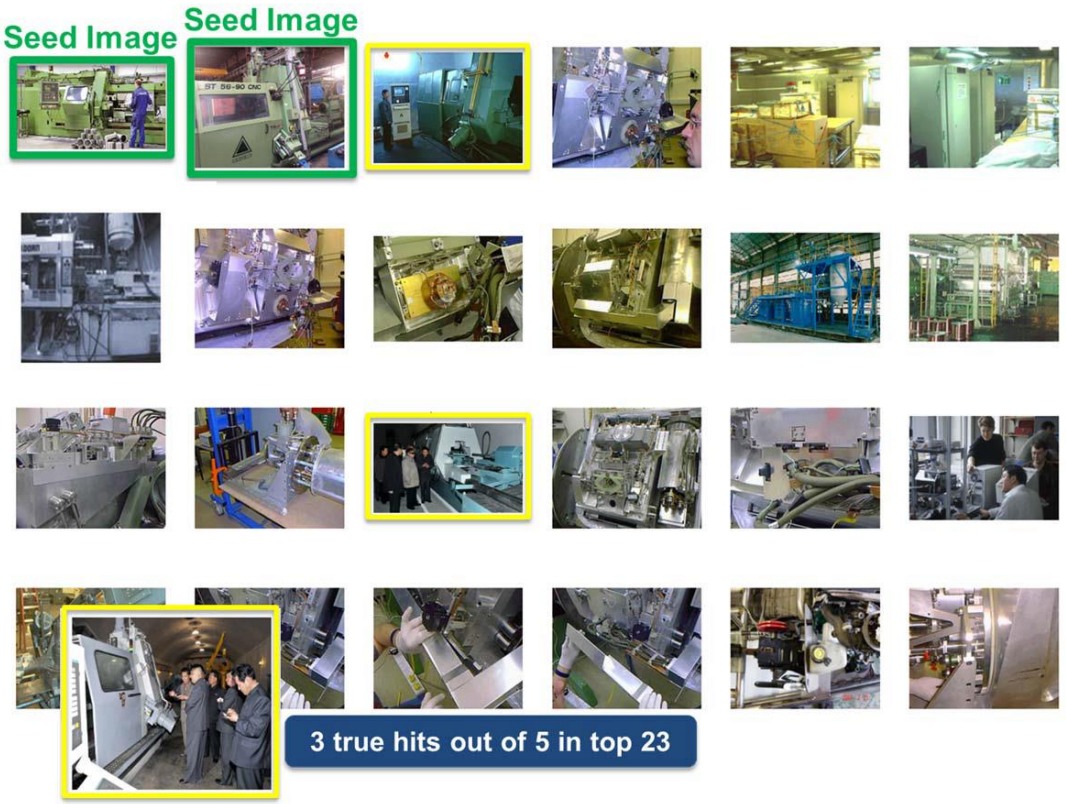
Image-to-Image



Retrieval Categories	Average Precision: 1 Seed Image	Average Precision: 5 Seed Images
Centrifuge	0.13	0.29
Centrifuge cascade	0.16	0.35
Cooling pond	0.46	0.58
Cooling tower	0.57	0.63
Hot cell	0.37	0.51
PWR	0.06	0.14
Magnox fuel element	0.02	0.03
Flow forming machine	0.34	0.46
UF_6 cylinder	0.32	0.5
Centrifuge schematic	0.09	0.14
Reactor core fuel matrix	0.27	0.47
CANDU fuel bundle	0.22	0.4
Mean Average Precision (mAP)	0.25	0.38

Approach: Embed NFC-related images in background set from ESP-Games dataset and assess how well the system can retrieve them

Stock Image-to-Real Image

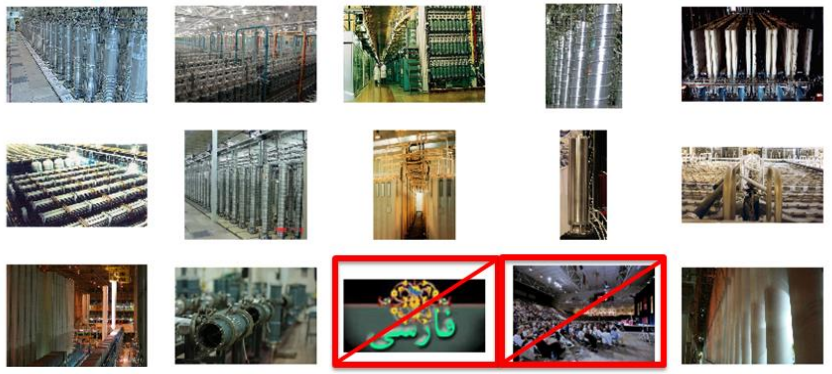


Approach: Embed NFC-related images in background set from ESP-Games dataset and assess how well the system can retrieve them

Text-to-Image

Retrieval Categories	Average Precision: Zero Shot	Average Precision: Few Shot
Centrifuge	0.0011	0.55
Centrifuge cascade	0.0006	0.49
Cooling pond	0.0017	0.80
Cooling tower	0.1526	0.91
Hot cell	0.0012	0.72
PWR	0.0011	0.37
Magnox fuel element	0.0002	0.11
Flow forming machine	0.0003	0.37
UF_6 cylinder	0.0062	0.89
Centrifuge schematic	0.0020	0.42
Reactor core fuel matrix	0.0006	1.00
CANDU fuel bundle	0.0016	0.66
Mean Average Precision (mAP)	0.0141	0.61

Text seed:
'centrifuge cascade'



Text seed:
'centrifuge', 'schematic'

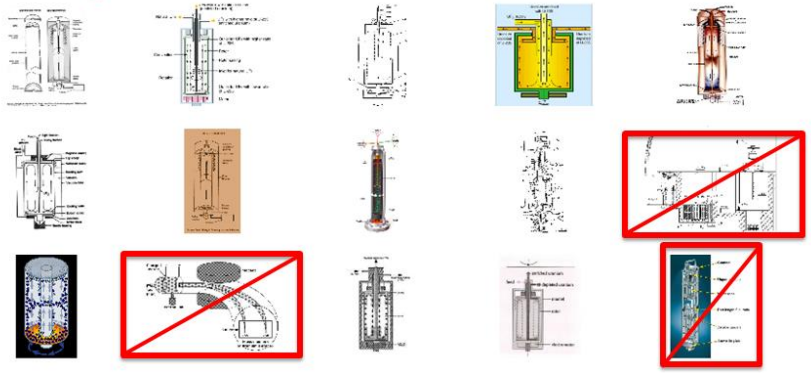
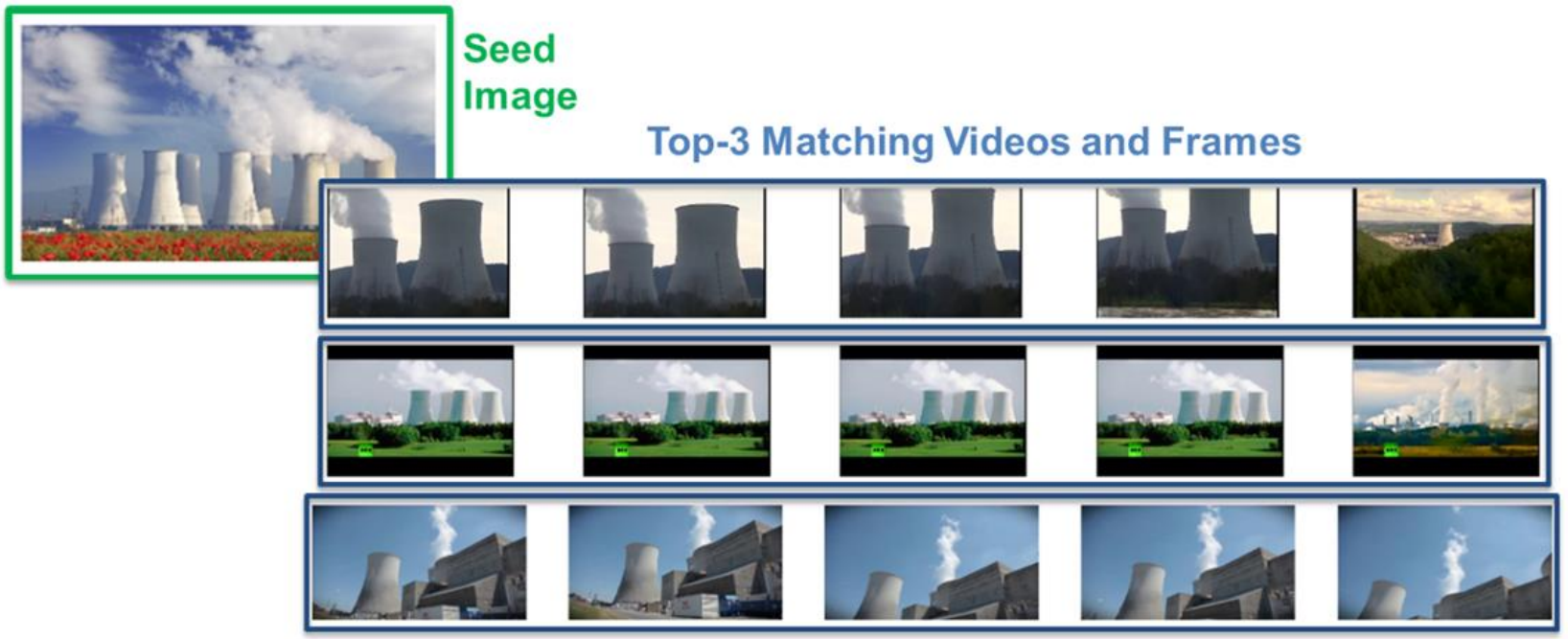


Image-to-Video



Approach: Find scenes in a collection of videos by performing frame-by-frame matching in the multimodal feature space with a seed image

Future Research

- Grow curated datasets in depth and breadth of NFC coverage
- Develop and incorporate video-specific DNNs that can focus on and follow objects
- Expand modalities to include transactional data (data collection and curation in progress)