SAND2019-1461PE

Detecting Anomalies in Safeguards Surveillance Data

PRESENTED BY

Zoe Gastelum, Michael R. Smith, Michael Hamel

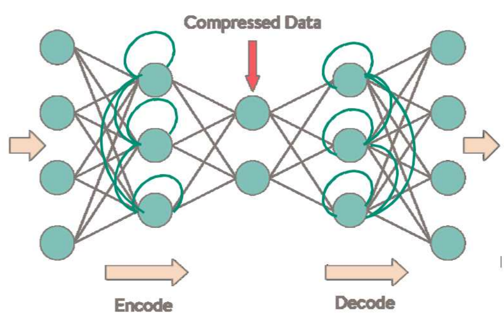
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

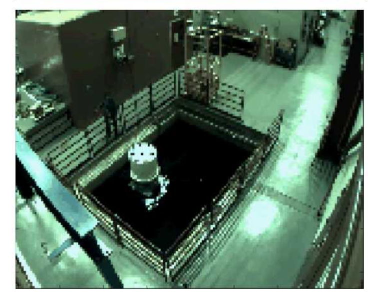
Overview

- **Motivation**: Reduce the IAEA surveillance review burden for inspectors by automatically identifying anomalies in image sequences
  - Distinct facilities, camera placement, picture taking interval (PTI)
  - Complements other capabilities such as object detection that are currently being developed
- **Proposal**: Evaluate unsupervised sequential learning to detect anomalous behaviors within surveillance data
- **Future outlook**: Use in combination with object detection capability such as YOLO
  - Detect anomalies around objects of interest
  - Share portions of the model to help enhance the training for each

Deep Predictive Coding Network (PredNet) Approach

- Unsupervised Learning
  - Trains on surveillance data without the requirement for manual human labeling or annotation
  - Learns "normal" behavior for each unique camera location
- Deep Convolutional Recurrent Network
  - Trains a model to predict next n video frames
  - Models spatial/temporal information





PredNet Approach (cont.)

- Use Conv-LSTM autoencoder to generate / predict the next frame given the previous frames
- Subtract the actual frame
- Compare the differences to detect anomalies
  - Threshold indicates spatiotemporal anomaly
  - Tune for false positives vs. false negatives
- Current example detects spatial anomalies (car) as well as temporal anomalies (pedestrians walking wrong direction)

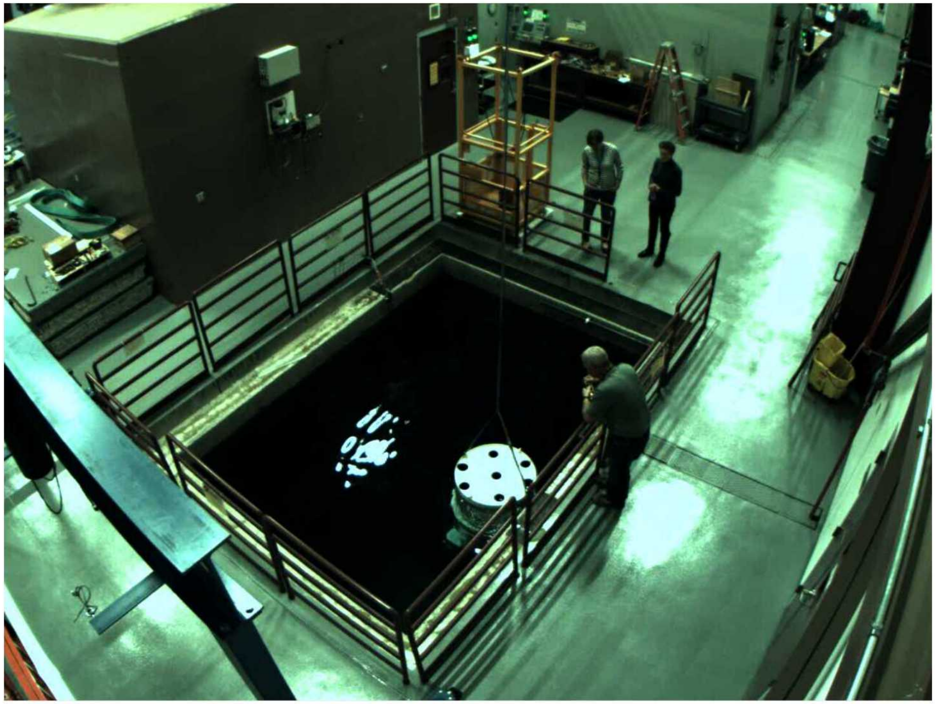
Image: Medel, J. R., & Savakis, A. (2016). Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390.

Why PredNet?

- PredNet is intended to detect anomalies in "patterns of life"
- As an unsupervised method, PredNet does not require human labeling or annotation of images
- By detecting "anomalies" rather than defined proliferation scenarios, we are not bound by known or expected off-normal scenarios
- We would like to investigate the applicability of PredNet to detecting safeguards-relevant anomalies. Other unsupervised methods such as image clustering could also be relevant.
- Disadvantage: because we are not manually defining anomalies, we do not directly control what it is flagging as anomalous.
  - PredNet anomalies might not be the same as safeguards anomalies
  - We can mitigate this through selection of training data to define "normal"
- PredNet was developed for video data, but we think it will work for protracted picture-taking intervals

Data

- Sandia deployed two NGSS cameras in the Gamma Irradiation Facility
- Collected down-time data and active scripted container movements over multiple days
- Collections include both full and empty floor vault scenarios



Scenarios for Data Analysis Plan

- We will evaluate what the PredNet algorithm determines as "anomalous" and its relevance to safeguards
- We will test four categories of potentially anomalous scenarios:
  1. **Unintentional Anomalies** – examine anomalies that are identified in "normal" operational scenarios
  2. **Intentional Anomalies** – intentionally insert anomalous frames to determine algorithm response
  3. **Operational Anomalies** – change operational activities within a facility, including types of containers present, appearance of containers, areas in which container are located
  4. **Safeguards scenarios** – experiment with scenarios that are determined to be of high safeguards interest, e.g. greyscale images, longer time lapse, and play-back loops

Results

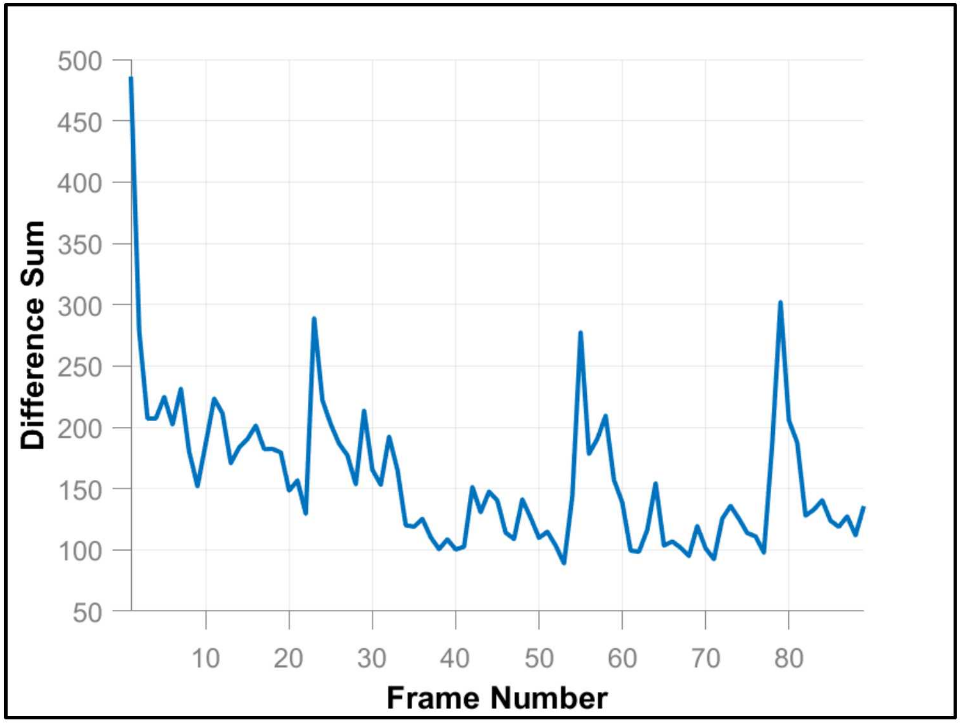
- Traditional display of results has been a visualization of anomaly scores overlaid on actual images
- Sandia is also exploring methods to quantify and compare anomaly scores resulting from each image
  - Each pixel in an image has an anomaly (or "error") score that quantifies the difference between actual and predicted images
  - We can use these scores to visualize and interpret anomaly scores over long temporal series of images to prioritize viewing individual frames

Results – Anomaly Overlay

A heat map visualization of anomaly scores is overlaid on actual images to indicate which pixel predictions we most different from the actual image.

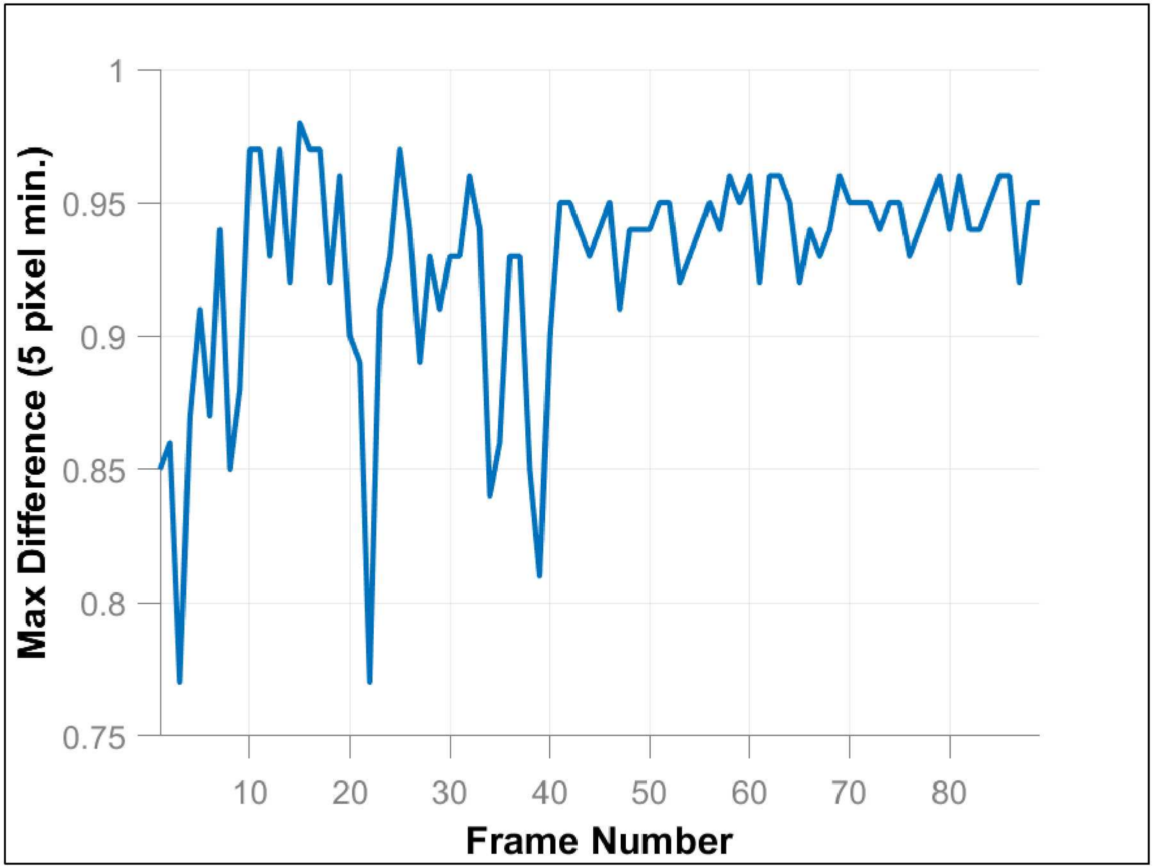
Results – Anomaly Sum per Frame

- The anomaly scores are summed over the entire frame
- Frame 1: starting point has high anomaly due to limited history from which to predict
- Frames 22/23: Container being lowered into the pool, people are entering the scene
- Frame 55: Crane component enters the scene
- Frame 80: Strobe light flashes and reflects



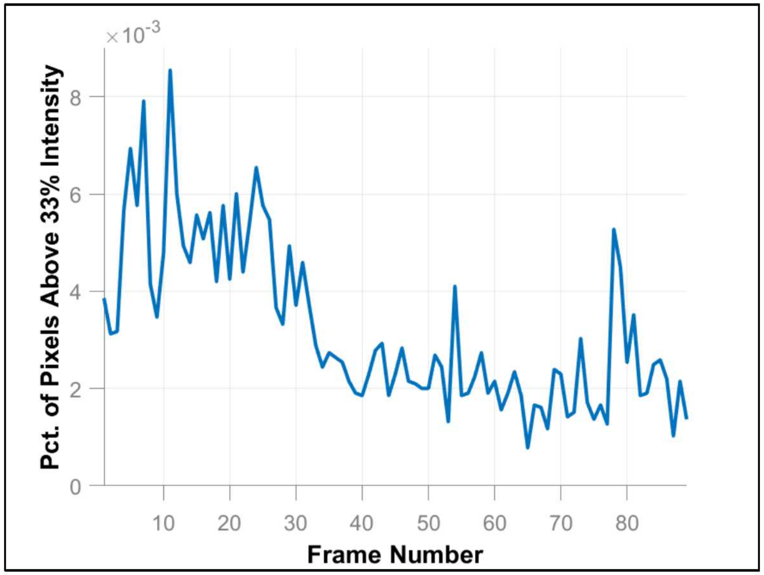
Results – Maximum Difference

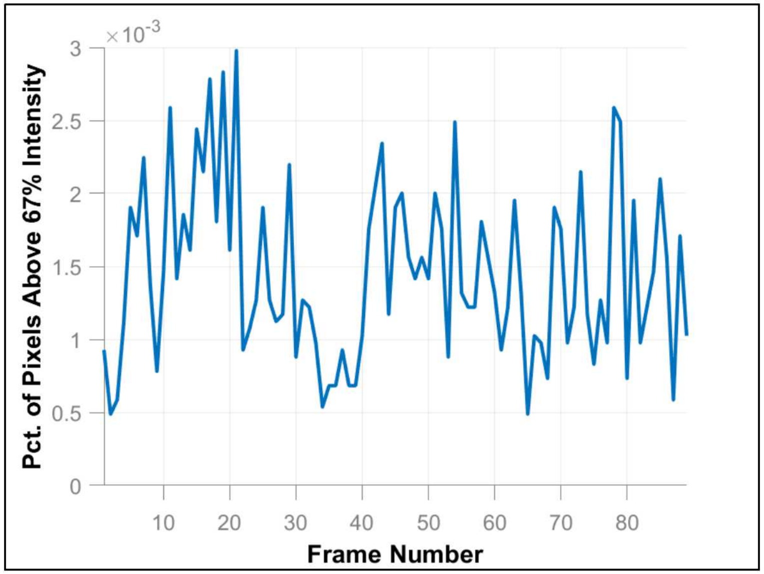
- The highest anomaly score in an image that includes five or more pixels is used to determined a maximum difference score
- Frames 10 – 20: Container is moving into the water, water ripples
- Frames 20 – 30: A group of people gather to watch the container go into the pool
- Frame 32: Crowd disperses



Results – Minimum threshold anomalies

- A minimum threshold for anomaly score is defined for the dataset, and the pixels in an image exceeding that threshold are counted
- We are determining which thresholds might provide the most useful visualizations, and how this contrasts with other methods
- Frames 10 – 20: Container is moving into the water, water ripples
- Frames 20 – 30: A group of people gather to watch the container go into the pool
- Frame 55: Crane component enters the scene
- Frame 80: Strobe light flashes and reflects





Discussion

- Questions or concerns?
- Do you see this as relevant for safeguards surveillance review?
- How could an anomaly detection capability like PredNet be used for safeguards?

Discussion

- What type of anomalous activity visible from surveillance cameras would be of highest interest for safeguards?
- If anomalous images can be identified in surveillance data, what is the preferred mechanism for highlighting them? Image overlays? Side-by-side comparison? Graphs of anomaly scores?
- What is the priority of loop detection? What length of looped footage would be realistic?