Proceedings of the INMM & ESARDA Joint Annual Meeting

Compression-based Analytics for Efficiently Identifying Events that Deviate from Standard Operating Procedures in Surveillance Video

Michael R. Smith
Sandia National Laboratories

Renee Gooding
Sandia National Laboratories

Christina L. Ting
Sandia National Laboratories

Jonathan Bisila
Sandia National Laboratories

ABSTRACT

Reviewing surveillance videos from nuclear safeguards facilities is a manually-intensive, tedious effort undertaken by nuclear safeguards inspectors who must identify any malicious activity or deviations from established protocols. While deep learning methods have demonstrated the potential for reducing the manual overhead, they require significant computational resources and specialized hardware. Inspectors, however, must often work on-site with the data and are frequently constrained to laptop computers. To meet the lower computational requirements, we extend *compression-based analytics* for efficient and effective spatial-temporal anomaly detection in video. Compression-based analytics are a class of machine learning algorithms that utilize data compression algorithms. At a high level, data compression algorithms aim to encode data in fewer bits than the original representation by *learning and removing* statistical redundancy. In the case of video anomaly detection, compression algorithms are used to learn patterns of standard operating activity. Normal operating activity will compress well and events that deviate from the standard operating behavior (i.e., anomalies) will not compress as well. On a variety of surveillance video data sets, we show that our method is competitive with state-of-the-art deep learning methods while requiring only a fraction of the computational resources. By including methods such as compression-base analytics into the Next Generation Surveillance Review tool, we seek to reduce the large burden placed on nuclear safeguards inspectors reviewing surveillance videos.

INTRODUCTION

The International Atomic Energy Agency (IAEA) Department of Safeguards is charged with monitoring the activities occurring in safeguarded nuclear facilities worldwide. Video surveillance is a core component employed by the IAEA to help verify the correctness and completeness of a State's declared nuclear-related activities at such facilities. Video surveillance footage allows inspectors to identify anomalous activity that deviates from standard operating procedures. In this paper, we define anomaly as anything statistically rare compared to events in the training data. The IAEA has historically relied on basic capabilities included in the General Advanced Review Software (GARS) to make the review process more efficient. Despite these capabilities, reviewing surveillance video remains a tedious task that requires significant manual effort and specialized expertise. The Next Generation Surveillance Review (NGSR) tool, which will replace GARS, provides several improvements, such as the inclusion of advanced machine learning (ML) and deep learning (DL) algorithms [1].

The opportunity to include ML and DL modules in the NGSR tool can reduce the large manual burden placed on inspectors performing surveillance review. Many proposed modules use DL

methods [2, 3, 4, 5, 6, 7], which have demonstrated better than human performance in computer vision tasks [8, 9] such as anomaly detection in video [10, 11]. Despite their success, DL methods require specialized hardware [12] and large amounts of data [13] to train and run efficiently. Unfortunately, many nuclear facilities do not have specialized hardware and the sensitive nature of the surveillance video precludes the data from leaving the facility. To overcome the challenges of DL, we propose to leverage compression-based analytics (CA), which are based on using data compression algorithms to compute a similarity metric between two data items, referred to as the normalized compression distance (NCD) [14]. CA has been successfully applied for a variety of ML tasks (e.g., classification, clustering, and anomaly detection) across diverse datatypes (e.g., genomic, language, and cyber) [15, 16]. Here we apply CA to identify anomalous activities that deviate from normal activities captured in a baseline set of training videos. In contrast to previous CA methods that focus on sequential representations of 1D tokens (e.g., DNA, natural text, and malware binaries) we extend CA to surveillance video, which can be considered a temporally ordered sequential representation of 2D images.

The remainder of this paper is organized as follows. We first provide a background discussion on the connections between the goals of data compression algorithms and DL, which motivate the development of CA. Our method, which we call *Spatial-Temporal NGram PPM (STNG PPM)*, for extending CA to surveillance video is described next. We then summarize the three data sets used in our study, two open-source data sets and one safeguards-specific data set from Sandia National Laboratories. This is followed by a presentation of our results. Across multiple benchmark data sets, we show that CA is competitive with the best DL methods. We conclude with directions for future work.

BACKGROUND

At a high level, DL algorithms seek to learn a lower dimensional representation of the data to learn the statistics of the data rather than memorizing the data; this can be viewed as a form of data compression [17, 18] [19]. As an example, autoencoders [17] are a class of unsupervised DL algorithms where there are no explicit labels that learn to reconstruct input data. Similar to a data compression algorithm, autoencoders are composed of an encoder and a decoder portion with a bottleneck module that is lower-dimensional than the input size. Intuitively, the bottleneck module represents the compressed knowledge that the network learns, which is analogous to the compressed representation that a data compression algorithm learns.

In supervised learning, where labels are provided, the same principle applies. There is generally a layer within a neural network that has fewer dimensions than the input. However, rather than reconstructing the input, the task is to predict which class an input belongs to. Reducing the dimensionality of the data forces the network to learn the statistics or patterns of the data that are associated with a particular class. Similarly, a data compression algorithm that seeks to learn statistical redundancies for better compression can be used to learn patterns in data that are associated with a particular class. This has implications for applying data compression algorithms for use in CA as the class with the data compression model that best compresses a new data item can be used for label prediction.

A key distinguishing factor between data compression algorithms and DL is the objective of each family of algorithms. DL seeks to minimize an error function, often the reconstruction error in the case of an autoencoder, and the prediction accuracy for supervised learning. In contrast, compression algorithms seek to minimize the statistical redundancy and the number of bits

required to transmit the information contained in the signal. Both compression and DL algorithms operate on the raw input data, however, DL methods can learn a feature transformation that is relevant to the task; this is one of the properties that makes DL extremely powerful, producing state-of-the-art results in many domains. However, as discussed in the introduction, DL methods have several requirements that make it challenging to apply them to surveillance video from nuclear facilities. In the next section, we introduce CA as an alternative to DL.

METHODS

CA refers to a class of ML techniques that leverage data compression algorithms to compute a similarity metric, known as the normalized compression distance (NCD), between two data items x and y. The NCD is defined by Li et al. [14]:

$$d(x,y) = \frac{|C(xy)| - \min(|C(x)|, |C(y)|)}{\max(|C(x)|, |C(y)|)},$$

where $|C(*)|$ denotes the compressed size of $*$ after applying a compression algorithm C, and xy denotes x and y concatenated. If x and y are very similar, then the compressed size of the two items concatenated is close to the compressed size of each item alone and $d(x,y) \approx 0$; if x and y are very dissimilar, then the compressed size of the two items concatenated is close to the sum of the compressed size of each item alone and $d(x,y) \approx 1$.

In the above equation, any real-world compressor can be used to obtain $|C(*)|$. Prediction by Partial Matching (PPM) with Arithmetic Coding (AC) [20] is an adaptive statistical data compression algorithm based on context modeling that is commonly used for calculating the NCD. Importantly, *PPM-AC* allows us to extend the application of data compression algorithms beyond their use in computing a similarity metric between two data items. Instead of compressing a data item using an adaptive context model of the current data item, $C(x)$, we can compress a data item using a fixed, previously-trained context model, $C(x|M)$, where M is the context model. How well (or poorly) data item x compresses using the context model M is a measure of how normal (or anomalous) that data item is with respect to what was observed in the data used to train M. As an example, if M is trained on surveillance video of standard operations, activities not observed during these periods will not compress well and will be identified as anomalies.

In our work, we use a modification of PPM-AC, called *NGram PPM* [21]. Instead of applying AC to do the actual compression to obtain $|C(x|M)|$, NGram PPM uses the PPM context model directly to compute a score $s(x|M)$ that indicates how well an item would be compressed. By skipping the actual compression step, NGram PPM is significantly faster than PPM-AC while providing equivalent results.

The original application of context modelling using PPM was intended for sequences of 1D tokens such as text. Here, we extend NGram PPM for application to video, which is a temporally ordered sequence of 2D frames. Our approach converts the 2D problem into a 1D problem in the following way. Each *region* in the video can be represented by a temporally ordered 1D sequence of values $x_i = z_{i1}, z_{i2}, \ldots, z_{in}$ for regions $i = 1, \ldots, m$, where n is the number of frames; see Figure 1. We note that a region can be defined for different spatial resolutions. At the highest resolution, a region may correspond to a single pixel; at a lower resolution, it is possible to group pixels together to reduce noise. Additionally, we explored the use of motion vector angles and magnitudes from motion estimation algorithms [22] as features rather than raw pixel values. Motion vectors describe

how pixels change from one frame to another. Importantly, the representation used to construct the 1D sequences constrains the type of anomaly identified. For example, RGB pixel magnitudes, motion vector angles, or motion vector magnitudes can be used to identify anomalies according to color, direction, or speed. Future work will involve other representations as well as combining information from multiple representations.

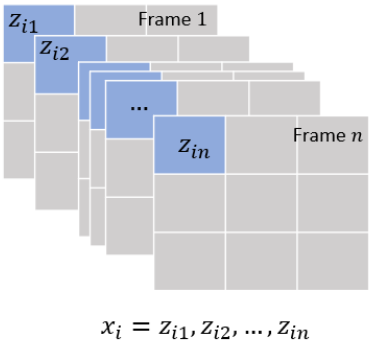

Figure 1: Extract temporally ordered 1D sequences from local regions of the video.

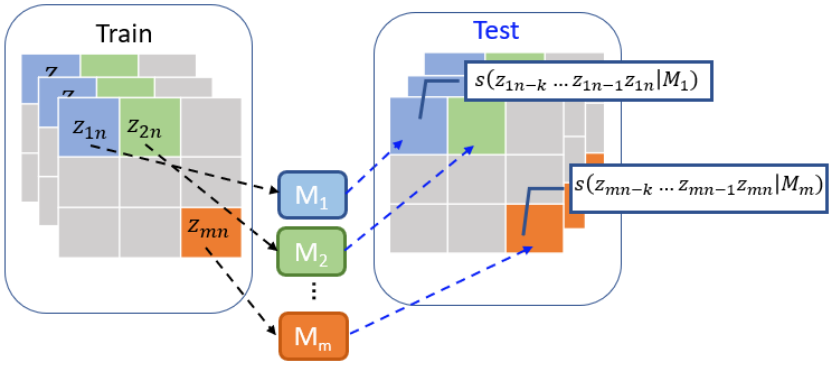

Figure 2: Model training and evaluation paradigm

The train and evaluate paradigm is depicted by Figure 2. Using video of baseline operations, our approach is to train local NGram PPM context models $M_1, M_2, \ldots, M_m$ using the 1D sequences for each region of the training video. New videos are evaluated by scoring each region using its corresponding local NGram PPM model. It is possible to assign a single score to each region of the video by using the entire sequence x_i. That is, it is possible to compute $s_i \equiv s(x_i|M_i)$. However, this score does not provide any temporal information about when the anomaly occurs. Instead, we assign a score for each region *of each frame*. Specifically, for each region $i = 1, \ldots, m$ of each frame $j = 1, \ldots, n$, we compute a score $s_{ij} \equiv s(z_{ij-k} \ldots z_{ij-1} z_{ij} | M_i)$, which indicates how well the NGram PPM context model M_i compresses a window of the sequence x_i, from $j - k$ to j, where k denotes the temporal window size. Regions and frames (i.e., times) with large scores will be flagged as anomalies. We refer to the approach of spatial and temporal anomaly detection as *Spatial-Temporal NGram PPM (STNG PPM)*.

We note that the local NGram PPM context models $M_1, M_2, \ldots, M_m$ are trained independently for each region and that this lends itself to parallelization. The number of regions m can also be

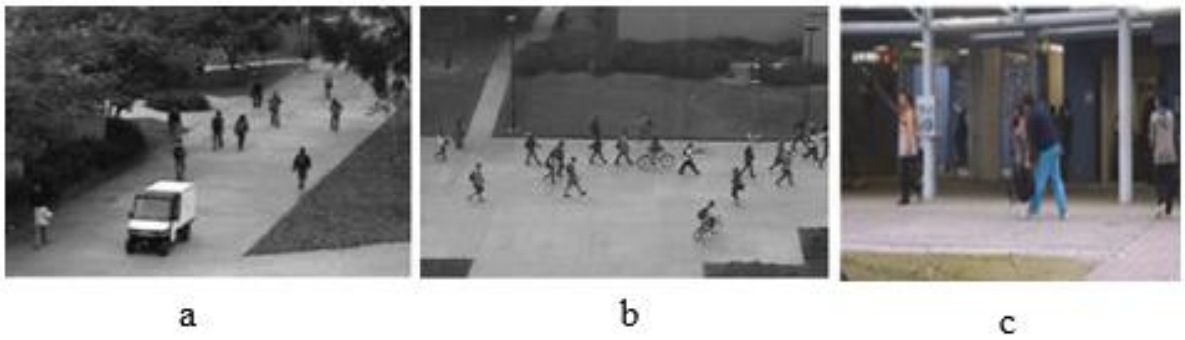

Figure 3: Example frames from the UCSD Pedestrians data sets (a) Peds1, (b) Peds2, and (c) the CUHK Avenue data set.

adjusted according to available computational resources and/or runtime demands. In future work, we will extend our approach to account for correlations between regions as well as longer-time dependent relationships.

DATA SETS

This section describes the data sets that we use to evaluate the performance of STNG PPM for anomaly detection in surveillance video.

UCSD Pedestrians Data Set

The UCSD Pedestrians data set [23] is a video data set acquired from a mounted stationary camera overlooking pedestrian walkways on the campus of the University of California San Diego (UCSD). Anomalies that are not seen in the training data include non-pedestrian activities such as people riding bicycles or driving small carts, as well as motion patterns such as pedestrians walking outside the flow of normal traffic. All anomalies are naturally occurring, that is, not staged. There are two data sets taken from different perspectives: Peds1 captures people walking towards and away from the camera; Peds2 captures movement parallel to the camera plane. Examples of frames from Peds1 and Peds2 are shown in Figure 3a and 3b respectively.

CUHK Avenue Data Set

The CUHK Avenue data set [24] was acquired at the Chinese University of Hong Kong (CUHK) campus. It contains video footage of a walkway and entrances to a building. Compared to the UCSD Pedestrians data sets, it is a more challenging including such as a slight camera shake and a larger variety of anomalies including novel objects not observed in the training video and staged anomalous events. A frame from this data set is provided in Figure 3c.

Sandia National Laboratories' Gamma Irradiation Facility

As a proof-of-concept use case for nuclear safeguards, we also examine data collected from Sandia National Laboratories' Gamma Irradiation Facility (GIF). Like many nuclear facilities, it contains a drying location and a spent fuel pool. Figure 4 shows the monitored area and the path use to bring a nuclear container to and from the drying area. The training video contains only video of containers exiting the drying area while the test video includes the containers both entering and exiting the drying area. There are several additional on-going activities in both the training and test videos, including a flashing light, people moving in the background, and a

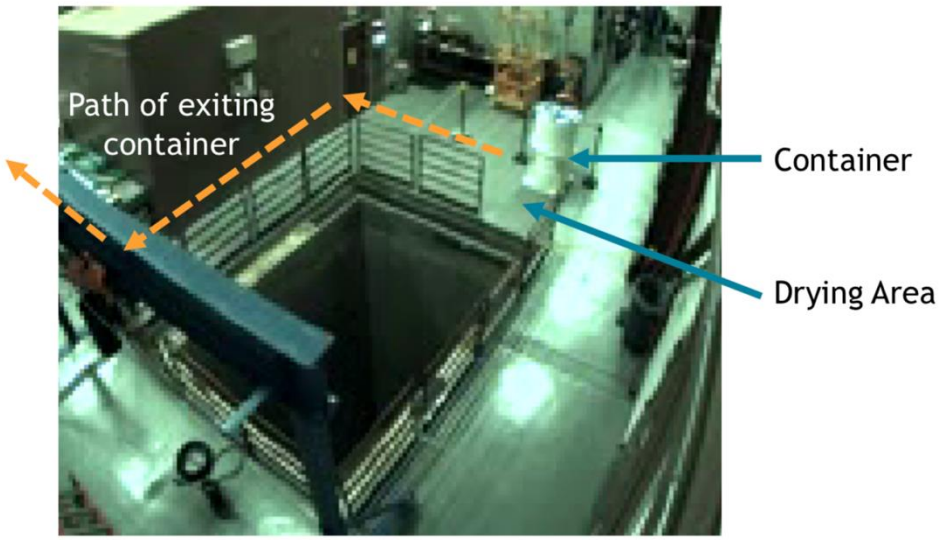

Figure 4. Example frame of the video from the Gamma Irradiation Facility at Sandia National Laboratories used to evaluate our workflow and the path followed for exiting containers. The path for containers entering the drying area follows the same path in reverse.

container being lifted and lowered into the spent fuel pool. In this scenario, we consider containers leaving the drying area as normal and containers entering the drying area as anomalous.

RESULTS

In this section we first evaluate STNG PPM on the previously described open-source data sets and show that it is competitive with state-of-the-art DL methods. We also evaluate STNG PPM on a data set collected at Sandia National Laboratories, which represents a more realistic setting for the nuclear safeguards application. All our computations take place on CPUs. We use Dual Socket Intel E5-2683v3 2.00GHz CPUs with 28 total cores. Since our algorithm is parallelizable, multiple cores speed up the computation but are not required.

To compare with DL methods, we adopt the frame-level criteria for calculating area under the curve (AUC). A frame is considered anomalous if *any* region in the frame is marked as anomalous. This criterion measures the performance of identifying anomalies in each frame and is commonly used in previous studies [23, 25]. However, there are multiple issues with using this evaluation measure. First, the frame-level AUC does not account for the localization of the anomaly. Therefore, if a frame contains an anomaly but there is no spatial overlap between the ground truth label and the algorithm prediction, a frame is still considered a true positive. This inflates the true positive rate. Second, if a frame does not contain an anomaly but even a single region in the frame is marked as anomalous (e.g., due to noise from camera shake), a frame is considered a false positive. This inflates the false positive rate. Alternative evaluation measures have been proposed to address these weaknesses [25] but have not been included in many prior works making it difficult to compare prior works with STNG PPM. Additionally, these new evaluation measures suffer their own issues. A detailed discussion is outside the scope of this paper and is being addressed as part of future work and is forthcoming.

The results of STNG PPM compared with the reported results from state-of-the-art DL methods are shown in Table 1. As discussed in the methods section, there are multiple ways to represent a region in STNG PPM. Because we are most interested in anomalous activities as opposed to objects, we choose to use motion vectors computed from an 6 x 6 block of pixels to represent a region [26]. That is, $x_i = z_{i1}, z_{i2}, \dots, z_{in}$ represents a 1D sequence of motion vectors in region $i = 1, \dots, m$, where m is much smaller than the original number of pixels. Thus, grouping pixels into blocks not only reduces noise, it also reduces the demand for computational resources, as each region can be computed in parallel. While the block size sets the spatial resolution, the temporal resolution is set by the window size $k = 3$ of the 1D sequence of motion vectors over which the STNG PPM scores are computed; recall the score is defined by $s_{ij} \equiv s(z_{ij-k} \dots z_{ij-1} z_{ij} | M_i)$. Additionally, we follow the procedure in many DL methods to smooth out the spatial and temporal noise. Specifically, we use a median filter across 3x3 regions and a median filter across 3 time steps. In calculating the frame-level metric, we also follow the method by Georgescu et al. [27], assigning the maximum NGram PPM score for each frame and then applying a Gaussian filter to the scores.

Table 1: Frame-level area under the curve (AUC) for a number of DL methods and our proposed CA method, STNG PPM.

Method	UCSD PEDS1 AUC	UCSD PEDS2 AUC	CUHK Avenue
Conv-AE [28]	81.0%	90.0%	70.2%
Conv-WTA-AE [29]	91.9%	92.8%	82.1%
GAN (ICIP 2017) [10]	**97.4%**	93.5%	---
Future Frame Prediction [30]	83.1%	95.4%	85.1%
Object-centric auto-encoder [11]	---	97.8%	**90.4%**
Appearance-motion cGAN [31]	---	96.2%	86.9%
$MLAD_{0+3}$ [32]	82.3%	**99.2%**	71.5%
Memory-Augmented AE [33]	---	94.1%	83.3%
Siamese Distance Learning [34]	86.0%	94.0%	87.2%
STNG PPM (ours)	**84.5%**	**90.5%**	**76.3%**

For UCSD Peds1 and Peds2, our method achieves favorable results compared to previous works using DL methods. Although STNG PPM does not outperform all the previous DL methods, we

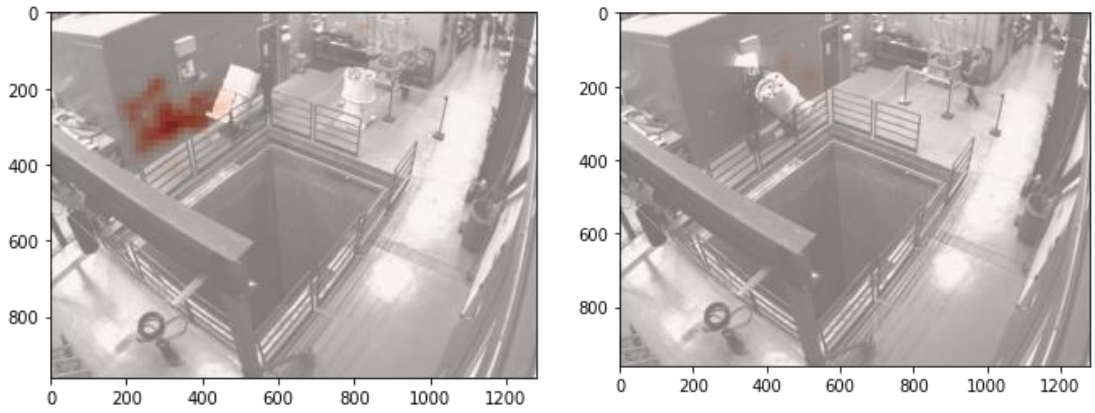

Figure 5: Heat map of STNG PPM scores for the GIF video. (Left) Anomalous activity of container entering the facility. (Right) STNG PPM scores of standard behavior of a container leaving the facility

emphasize that it does not require the use of GPUs and instead takes advantage of multiple CPU cores for parallel execution. For the CUHK Avenue data set, STNG PPM achieves a frame AUC score of 76.3%, which is lower than the DL methods. Although STNG PPM is not restricted to using only motion information, our current implementation operates on only the motion vectors limiting anomaly detection to only anomalous movement. Thus, novel static objects are not detected. If we remove the first two CUHK videos in which a non-moving bag present in the foreground is considered anomalous, STNG PPM achieves a frame AUC of 84.0%, which is more competitive with state-of-the-art DL methods.

For the GIF video captured at Sandia National Laboratories, we marked each frame in the test video as anomalous if a container was moving into the drying area. Our method achieves a frame AUC of 94.5%. Frames overlaid with the scores from STNG PPM for a container being moved in (anomalous, left) and for a container being moved out (normal, right) are shown in Figure 5. Anomalies are indicated by higher scores, or higher color intensities, in the heat map. We observe that the frame with a container being moved in (left) has higher scores around the container, as compared with the frame with a container being moved out (right). Clearly, this differentiates the anomalous activity (spatially and temporally) from standard operating conditions.

CONCLUSIONS AND FUTURE WORK

We have presented a workflow leveraging compression-based analytics (CA) to identify anomalous activities in video. On open-source data sets as well as on video from a safeguards-relevant facility, we demonstrated that our method, *Spatial-Temporal NGram PPM (STNG PPM)*, is competitive with state-of-the-art DL methods without requiring specialized hardware. The opportunity to include CA into the Next Generation Surveillance Review (NGSR) tool can help alleviate the tediousness of reviewing surveillance video and reduce fatigue-induced errors.

There are several key areas for future work to improve on STNG PPM. Specifically, we will consider 1) fusing other video representations with the motion vectors to detect more complex anomalies (e.g., based on both object *and* motion), 2) incorporating spatial dependencies between regions when training the local models, and 3) applying time series analysis techniques to discover anomalous patterns and trends. From a more practical perspective, future work also needs to

examine the applicability of STNG PPM to safeguard specific datasets and improved engineering for scalability.

ACKNOWLEDGEMENTS

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. 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. SAND2023-02849C

REFERENCES

[1] M. Thomas, A. A. Alessandrello, S. Rocchi, E. Galdoz, M. John, C. Brunhuber, M. Moeslinger, A. Smejkal, K. Ruuska and J. Pekkarinen, "Next Generation Surveillance Review (NGSR): Enhancing The Safeguards Toolkit," in *The Joint INMM-ESARDA Annual Meeting*, 2021.

[2] S.-H. Park, "Diversion Detection using Optical Surveillance based on Deep Learning," in *The Joint INMM-ESARDA Annual Meeting*, 2019.

[3] M. R. Smith, D. Hannasch, M. Hamel, M. Thomas and C. Gaitan-Cardenas, "A Deep Learning Workflow for Spatio-Temporal Anomaly Detection in NGSS Camera Data," in *The Joint INMM-ESARDA Annual Meeting*, 2021.

[4] M. Thomas, S. Passerini, Y. Cui, J. Rutkowski, S. Yoo, Y. Lin, J. H. Park, M. R. Smith and M. Moeslinger, "Deep Learning Tequnies to Increase Productivity of Safeguards Surveillance Review," in *The Joint Annual INMM-ESARDA Annual Meeting*, 2021.

[5] Y. Lin, X. Zhang, J. H. Park, S. Yoo, Y. Cui, M. Thomas and M. Moeslinger, "Using Machine Learning to Track Objects Across Cameras," in *The Joint INMM-ESARDA Annual Meeting*, 2021.

[6] Y. Yokochi, S. Chen and K. Demachi, "A Novel Detection Approach to Preventing Theft of Nuclear Materials Using Deep Learning-based Object Detection and Human Pose Estimation," in *The Joint INMM-ESARDA Annual Meeting*, 2021.

[7] E. Wolfart, A. Casado Coscolla and V. Sequeira, "Deep Learning for Video Surveillance Review," in *The Joint INMM-ESARDA Annual Meeting*, 2022.

[8] K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers:Surpassing Human-Level Performance on ImageNet Classification," in *Proceedings of the IEEE international conference on computer vision*, 2015.

[9] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *International Conference on Machine Learning*, 2015.

[10] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[11] R. T. Ionescu, F. S. Khan, M.-I. Georgescu and L. Shao, "Object-centric Auto-encoders and Dummy Anomalies for Abnormal Event Detection in Video," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[12] N. C. Thompson, K. Greenewald, K. Lee and G. F. Manso, "Deep Learning's Diminishing Returns," *IEEE Spectrum,* vol. 58, no. 10, pp. 50-55, 2021.

[13] J. Hestness, S. Narang, N. Ardalani, G. Diamos, H. Jun, H. Kianinejad, M. A. Patwary, Y. Yang and Y. Zhou, "Deep Learning Scaling is Predictable, Empirically," *arXiv preprint,* p. arXiv:1712.00409, 2017.

[14] M. Li, X. Chen, B. Ma and P. Vitanyi, "The Similarity Metric," *IEEE Transactions on Information Theory,* vol. 50, no. 12, pp. 3250-3264, 2004.

[15] R. Cilibrasi and P. M. B. Vitanyi, "Clustering by Compression," *IEEE Transactions on Information Theory,* vol. 51, no. 4, pp. 1523-1545, 2005.

[16] C. Ting, N. Johnson, U. Onunkwo and J. D. Tucker, "Faster classification using compression analytics," in *International Conference on Data Mining Workshops* , 2021.

[17] M. A. Kramer, "Autoassociative Neural Networks," *Computers & Chemical Engineering,* vol. 16, no. 4, pp. 313-328, 1992.

[18] G. E. Hinton, A. Krizhevsky and S. D. Wang, "Transforming auto-encoders," in *International Conference on Artificial Neural Networks*, Espoo, Finland, 2011.

[19] J. Balle, V. Laparra and E. P. Simoncelli, "End-to-end optimized image compression," in *International Conference on Learning Representations*, 2017.

[20] A. Moffat, "Implementing the ppm data compression scheme," *IEEE Transactions of Communications,* vol. 38, no. 11, pp. 1917-1921, 1990.

[21] T. Bauer, "NgramPPM: Compression Analytics without Compression," Sandia National Laboratories, 2021.

[22] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE transactions on Image Processing,* vol. 9, no. 2, pp. 287-290, 2000.

[23] W. Li, V. Mahadevan and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, no. 1, pp. 18-32, 2013.

[24] C. Lu, J. Shi and J. Jia, "Abnormal Event Detection at 150 FPS in Matlab," in *Proceedings of the IEEE international conference on computer vision*, 2013.

[25] B. Ramachandra and M. Jones, "Street scene: A new dataset and evaluation protocol for video anomaly detection.," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020.

[26] scikit-video developers, "scikit-video: video processing in Python," [Online]. Available: http://www.scikit-video.org/stable/. [Accessed 2023 04 2023].

[27] M. I. Georgescu, R. T. Ionescu, F. S. Khan, M. Popescu and M. Shah, "A Background-Agnostic Framework with Adversarial Training for Abnormal Event Detection in Video," *IEEE transactions on pattern analysis and machine intelligence,* vol. 44, no. 9, pp. 4505-4523, 2021.

[28] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury and L. S. Davis, "Learning Temporal Regularity in Video Sequences," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[29] H. T. Tran and D. Hogg, "Anomaly Detection using a Convolutional Winner-take-all Autoencoder," in *Proceedings of the British Machine Vision Conference*, 2017.

[30] W. Liu, W. Luo, D. Lian and S. Gao, "Future frame prediction for Anomaly Detection - A New Baseline," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[31] T.-N. Nguyen and J. Meurier, "Anomaly Detection in Video Sequence With Appearance-Motion Correspondence," in *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[32] H. Vu, T. D. Nguyen, T. Le, W. Luo and D. Phung, "Robust Anomaly Detection in Videos Using Multilevel Representations," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

[33] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh and A. van den Hengel, "Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection," in *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[34] B. Ramachandra, M. Jones and R. Vatavai, "Learning a distance function with a Siamese network to localize anomalies in videos," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020.