2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) | 978-1-7281-8708-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/AIKE48582.2020.00018

Knowledge Graphs for Semantic-Aware Anomaly Detection in Video

Alina Nesen
Purdue University
West Lafayette, USA
anesen@purdue.edu

Bharat Bhargava
Purdue University
West Lafayette, USA
bbshail@purdue.edu

***Abstract*—Video understanding, surveillance and analytics fields have been dynamically expanding over the recent years due to the enormous amount of CCTV, dashcams and phone cameras which generate video data stored on cloud servers, in social networks, in public and private repositories. The video data has a great potential to be used for improving situation awareness, prediction and prevention of unwanted events and disasters in various settings. Still, there is a significant need for methods and ways to understand the large amount of video recordings and to extract hidden patterns and knowledge. Deep learning networks have been successfully applied for video object and anomaly detection tasks. However, while neural networks focus on utilizing features within an object to be detected, the vast amount of background knowledge remains unnoticed. We propose a semantics centered method for video anomaly detection which allows to identify entities that are inconsistent with the scene and thus can be marked as a potential anomaly. Our method is inspired with the way humans comprehend the surroundings with incorporating external knowledge and previous experience. As a source of external knowledge for deep learning networks we maintain a knowledge graph which allows to compute semantic similarity between the detected objects. Similarity of the entities in the frame depends on the distance between the graph vertices which represent the recognized entities. The object which is semantically distinct from other entities in the video is an anomalous one. We conduct experiments on real-life data to empirically prove the efficiency of our approach and provide an enhanced framework that leads to anomaly detection in video with higher accuracy and better interpretability.**

***Keywords*—anomaly detection, semantic extraction, video understanding, knowledge graphs, surveillance systems**

I. Introduction

In recent years, along with enormous spread of video recording devices, social video platforms such as YouTube, TikTok, and an exponential growth of available storage space for the videos, there has been a significant progress in the development of the methods for video understanding, object detection, recognition and tracking, scene segmentation and activities understanding. Currently, most of these procedures are performed with pre-trained neural networks that require substantial labeled datasets for their training. Still, there is an apparent deficiency of knowledge extraction and analysis methods for video recordings and an unworkable ratio of cameras to human monitors. One crucial task in video analysis is timely detection of anomalous and outstanding events, objects and activities. Another important aim of a practical video analysis and an anomaly detection system is its ability to send a well-timed signal about an activity that deviates normal patterns. Once an anomaly is detected, it can further be categorized into one of the specific activities using classification techniques or human expert knowledge.

Anomaly detection in video and image data is not a trivial task since real-world anomalous events are complex and diverse. The space of all the possible anomalous events is extraordinarily vast and impossible to foresee in advance when constructing training and testing datasets for supervised learning. Therefore, it is highly desirable for the anomaly detection algorithm to be independent from the training dataset provided beforehand and to be conducted with minimum or no supervision.

The first step towards addressing anomaly detection is development of the algorithms that can be trained to identify a specific anomalous event, for example a deep learning network for violence detection or traffic accident detection. This is achieved by training on a comprehensive dataset that contains normal behavior of the system as well as anomalous frames, e.g. ones with car accident. However, such solutions do not generalize to catch each and every inconsistent and abnormal event, and most of the time they cannot be transferred between different systems, situations and data. They require re-training and oftentimes expensive manual labeling of new datasets The need for solutions that do generalize inspired the proposed method which does not require additional labeled datasets for recognizing anomalies but instead addresses the relationships between the entities detected with a neural network from the knowledge graph.

Deep learning methods have proved to demonstrate near human or better than human accuracy for video understanding in recent years [1]. Neural networks were initially inspired by human brain and followed the idea of getting an understanding about the world by propagating the information obtained from the input through a chain of interconnected functions, or neurons, which learn to adjust their weights and parameters in a way that they can output a correct final decision about an object or an event. Deep learning systems, however, do not exploit an important fact of human nature: when making a decision, people tend to rely on the knowledge they had acquired previously and not just on the vast amount of alike data that is available at the current moment. Humans do not depend exclusively on the representation of the immediately seen and perceived data but use external knowledge associated with the data which may signal about possible anomalies. This is the way a human being detects an anomaly: after she detects something that is not in line with the surroundings, she uses her experience and the knowledge she has accumulated up to the current moment to classify the anomaly and determine further course of actions. In the example provided in *Figure 1*, a human would easily sense that a yellow car stands out in the environment because of its price and model. For the machine, such external knowledge can also be attained in advance, prepared by an expert and stored in the manner that makes this knowledge easily accessible, evaluable and analyzable. In this paper, we explore the way of storing this knowledge in knowledge graphs.

978-1-7281-8708-2/20/$31.00 ©2020 IEEE
DOI 10.1109/AIKE48582.2020.00018

Knowledge graphs like the deep learning systems mimic how the human brain works but in their own unique and different way. They represent the knowledge domain and the relations between entities in it.

Knowledge graphs have become widespread for dealing with storing and organizing large volumes of information however they have not been examined for the ability to further extract and explain supposed anomalies identified by machine learning algorithms. Along with the knowledge graphs there can be other ways to store external knowledge in the ways that allow to calculate semantical distances between entities. These methods can be used separately or together in order gain awareness and understanding of the input video data.

A number of organizations in academia and industry invest in development of their own knowledge graphs and knowledge bases, such as Wikipedia, Freebase, YAGO, Microsoft's Satori, and Google's Knowledge Graph [2,3]. There are works that utilize knowledge graphs in video understanding, classification, recognition and captioning [4], however the enormous potential of the external knowledge for distinguishing possible anomalies in the video and image data was not yet used. With the abundancy of tools and methods to extract semantics from the videos the next logical step in the research would be to connect it with the knowledge base that has further descriptions, additional properties and relations defined between the recognized objects and scenes and connections of these objects with the external world. Establishing such links in the knowledge graph for identification abnormal objects or events in the video also helps to achieve another important goal which is providing an explanation for the decision making system.

Additional reason that makes anomaly detection a difficult task is the fact that the boundary between normal and anomalous behaviors is often indistinct, uncertain and not easy to identify. In real life, the same behavior could be considered normal or abnormal behavior depending on the context and conditions. While the research on making conventional neural networks explainable and take the stigma of black-box mechanisms off them is ongoing, our method provides interpretable ways to classify anomalies and even move the threshold which specifies the boundary between an anomalous and normal event.

Robust and reliable solutions for detecting anomalous events in surveillance video data are increasingly important in the disaster prevention context, in which failure to discover an anomaly may have a significant impact on the decision-making process and the result of the mission. Unmanned aerial vehicles consistently collect large amounts of video data which need to be processed and analyzed. Anomaly detection is one of the main areas where research is needed for development accurate and fast methods for video analysis [5].

Other possible scenarios that motivate our search for semantical anomaly detection are building safe communities during disasters and outlier events, such as floods, earthquakes, fires. We describe a case study from the surveillance cameras of the flooded city of Tbilisi in 2015 [6]. We collect and create the video dataset that contains anomalous events which could be found on the city streets during the time of the flood to use as a case study for our experiments. Besides disaster prevention, the suggested method can find applications in street surveillance, mission completion in military setting, agricultural and medical applications.

Inspired by the way humans learn, we aim to add background or external knowledge at each step of video processing to measure the probability of incoherent or inconsistent objects at any given time captured in the frame. Understanding of the semantics ("meaning", "context") of the data is crucial for this kind of inference. We employ semantic computing principles of data and knowledge engineering [7] for capturing this additional context and use existing semantic similarity methods and techniques from natural language processing to measure distance between ontologies and to quantify semantical similarity between the embedded vectors obtained from the knowledge captured from the video frames.

Our proposed approach takes advantage of the existing family of deep learning-based techniques for object recognition and of the concept of knowledge graphs. Deep learning networks such as YOLO [8] provide a way for speedy real-time object detection, which can be further improved with the help of the background knowledge that is obtained from the knowledge graphs. Knowledge graphs can represent a specific knowledge domain related to the entities detected in video data or be a common-sense graph and applied to a wide range of domains. The synergy of the two approaches for understanding of the contents of the video, namely deep learning and knowledge graphs, helps to capture new unexpected and abnormal occurrences of the objects in the data which would otherwise go unnoticed because of the absence of preliminary training of the neural networks for the specific anomalies.

For example, the surveillance camera for parking lot records video that can be analyzed with the neural network which learned to identify car models (Fig.1 displays camera footage frame with the models Lamborghini, BMW, Lada). However, the recognition system is oblivious to the relationships between these identified models and cannot indicate that supercar in the poor district is an anomalous event without previous expensive training since the system was trained on the objects from the same distribution. Expanding the datasets to include out-of-distribution (OOD) objects marked as anomalous is an expensive task and can never be performed taking into account all possible OOD objects since there is an infinite number of those. On top of that, the OOD training dataset would have to incorporate a vast amount of possible contexts where an event (object) detected can be anomalous or not depending on the context. E.g. supercar captured at the surveillance camera in the vicinity of the Top Marques Monaco should not be treated as an anomaly based on the context.

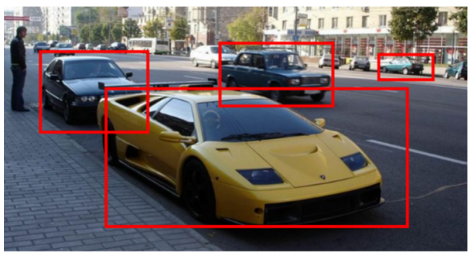

Figure 1 Object recognition in real-life scene can be trained to distinguish different car models

Our method follows the intuition of human approach in anomalies detection. Once the objects are identified, there is a

need for an external source of knowledge that can specify the relationship between the objects. In the example above related to the car models, a knowledge graph that describes the car classification clearly demonstrates the dissimilarity of one of the detected cars from the rest of the vehicles (Fig.2) since the Lamborghini car model belongs to the Supercars class while all other vehicles belong to the Compact cars class. By design, the knowledge graph contains compact cars and more expensive cars in different subgraphs. The distance between a car model in the *Supercars* class and a car in the *Compact car* class is greater than the distances between the cars within the *Compact car* class. The graph is undirected and unweighted which allows to take advantage of Seidel's algorithm for finding the shortest-paths lengths via matrix multiplication [9]. As a result of the computation, the video frame which contains cars from both *Supercars* and *Compact* classes will indicate larger dissimilarity than the frame where all classes fall into one subgraph of the knowledge graph, i.e. only *Compact* classes were detected.

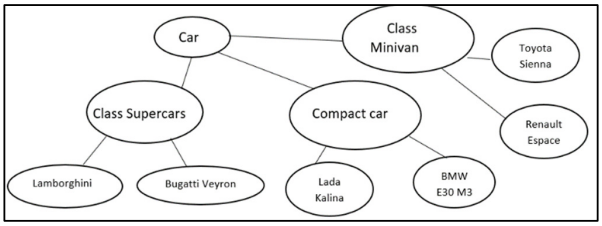

Figure 2 A tiny subgraph from the knowledge graph for vehicle classification domain

Thus, using the additional relationship between the detected objects which are stored in the knowledge graph, the proposed framework will distinguish an object that noticeably stands out among the rest. In the video fragment displayed on Figure 1, the luxury car model would draw attention of humans in the district but if the surveillance video analytics systems can involve external knowledge to mimic this approach then it can trigger a notification to a human expert to investigate a potential anomaly.

Another example which is described in detail in the section with the experiments analyzes video dataset from the city of Tbilisi during the 2015 flood when the wild animals escaped the zoo: bears, lions, tigers were found on the city streets and captured by the cameras (Fig.3). The relation between the ontologies of the knowledge graph and objects detected on the video is calculated with the similarity measures between the entities that are present in the frame: "car", "traffic sign" are semantically close to the "city", while "hippopotamus" is not, thus marking this frame as suspicious.

Figure 3. Screenshot from the video dataset of the Tbilisi city streets during the flood of 2012.

The proposed solution is aimed to enhance semantic understanding of the visual data in order to identify the anomalous content with providing explainable paths that were used for flagging the anomaly.

Our contributions are as follows:

1. We propose a new interpretable approach for detecting anomalous events in video data. Our method is based upon existing object recognition techniques and provides semantic understanding of the video. It further incorporates existing or newly constructed knowledge graphs as a basis for similarity computation.

2. We experimentally show that our method can identify anomalies that otherwise would not be recognized by the systems unless they had been specifically trained to learn to label such events as anomalous in the training dataset event collection.

3. To evaluate a use case with the semantical outliers from the video data, we collect the Tbilisi flood video dataset which contains recorded video with the anomalies (wild animals in the city streets). This dataset is a recording of the city streets video during the flood of 2015. It can serve as a benchmark for the methods of semantic-aware anomaly detection.

II. Related Work

In this section we review the current work in anomaly detection in video. The solution suggested in this paper involves actively researched methods of computer vision, object detection and recognition, scene understanding, semantic similarity methods and knowledge graphs. Hence, we review both the anomaly detection papers which concentrate on semantic meaning of the scenes as well as the latest advances in the applications of deep learning, knowledge graphs and semantic similarity computations in the computer vision setting.

In [10], the authors formulate the problem of detecting semantic anomalies as the out-of-distribution detection as well as provide a review of the current interest in OOD detection and classifying these objects as anomalous. They approach the problem with building a multi-task learning framework with auxiliary objectives.

Improving the neural network with knowledge graphs has been proposed in [30, 4] where the authors investigate the use of structured prior knowledge for improving performance of image classification tasks. The notion of semantic consistency is employed to quantify and generalize the knowledge and take into account background knowledge when assigning labels to the objects being classified.

To the best of our knowledge, there are no published methods that combine knowledge graphs with deep learning for the reverse problem, that is, for anomaly detection, where the background knowledge is leveraged to spot an inconsistency as opposed to stimulation the consistent labeling.

Anomalies in video surveillance systems have been extensively studied; the main methods used for unsupervised anomaly detection incorporate principal component analysis (PCA), autoencoders and their modifications in the form of convolutional autoencoders (CAEs), de-noising autoencoders (DAE), deep belief networks (DBN), long short-term memory

networks (LSTM) and their modifications, such as ConvLSTM, FC-LSTM. [13, 14, 15, 16, 17, 18]

In other domains, such as a military setting, surveillance is performed using fixed radar stations or patrol aircrafts, and it can help discover unlawful, unsafe, hostile, and anomalous behavior. The same applies to agricultural industry where large fields of crops are surveilled with the unmanned drones. These vehicles collect a large corpus of video data, and the main challenge persists: manual detection of such behavior is infeasible. Machine learning approaches are utilized to detect deviations from the normal models. An early approach to maritime anomaly detection use the neural networks that predict normal vessel speed taking into account such features as port location, current location and direction of travel [19, 20]. These are supervised methods that utilize extensive labeled datasets. However, obtaining annotations is a laborious task, especially for video data, that can be very time and budget consuming.

For video surveillance applications in the urban scenario, there are several attempts to detect violence or aggression in videos. Datta et al.[16] proposed to detect human violence by exploiting motion and limbs orientation of people. Kooij et al.[17] employed video and audio data to detect aggressive actions in surveillance videos. These methods are primarily expensive on the human labeling part. Moreover, all of them lack explainability and need additional research to add an interpretability component.

Deep neural networks are known as black box systems named so because of their un-explainability by humans when the result is obtained. Explainability is of crucial importance in modern deep learning systems for it may increase one or more of the following in the system: (1) Transparency and interpretability, (2) Effectiveness by helping a human expert to make an informative choice regarding the suggested classification without bias and unfairness, (3) Raise trust to AI [11]. The suggested framework, which is based on semantic meaning extraction, associating the meanings with natural language terms and finding relations between them has a higher degree of interpretability since humans also operate with natural language concepts for semantic understanding.

Understanding and explaining video is mostly achieved with the deep learning methods, namely LSTMs and deep recurrent neural networks (RNN) which now leading in the area of speech recognition, sequence modeling and image captioning which is closely related to video understanding [18]. These DNNs are used both for encoding (extracting the entities) the video as well as for the decoding (in other words, generating text in natural language that describes the video). YOLO, popular framework for object detection based on CNN architecture, emphasizes speed as the main motivator of its efficiency. It achieves the needed result with specifically designed loss function.

Video understanding can be made more comprehensive if a scene is detected along with the objects and entities. Several attempts have been made to model and learn a scene. In general, scene understanding expands awareness on the scene structure (e.g. pedestrian sidewalks, intersections, parks, eateries), scene status (e.g. flood, traffic jam), scene motion patterns, etc. With the knowledge of scene structure, activities and motion patterns, low-level tracking and abnormal activity detection (anomalous motion detection) can be improved.

Knowledge graphs have been extensively researched and used for information processing and organization for several decades. The first knowledge graphs were called ontologies or semantical graphs. With the rise of big data and Google search engine in particular, they have evolved into powerful graph databases that describe multiple attributes and relationship between entities. Knowledge graphs have been used for visual question answering and relationship extraction. Jia et al.[32] construct a heterogeneous relation network, a special form of knowledge graph, to capture anomalies from streaming multimodal data. For a comprehensive survey of graph-based anomaly detection, we refer readers to [33]. Graph neural networks and graph embeddings for anomaly detection were proposed in [34,35, 36].

The last step in anomaly detection in the proposed solution is the semantic similarity computation. There are a number of existing techniques in natural language processing that provide quantification mechanisms for semantical data: shallow neural networks such as word2vec, topic modeling methods such as LSI and LDA, TF-IDF [19, 20, 21]. The knowledge graph mining algorithms for quantifying semantic similarities between different nodes and entities in knowledge graphs are available and can be connected with the data extracted from heterogeneous sources via machine learning.

III. Framework Components

Deep neural networks for object detection

Neural networks for object detection are usually trained in a way that they can predict a bounded box around the object that is expressed through the spatial coordinates of its top-left corner and its width and height. YOLO model [8] divides the input image into an *SxS* grid and each grid cell is responsible for predicting the object that would have center within that cell. Each grid cell predicts *B* bounding boxes and their corresponding confidence scores. At the same run, independent of the number of boxes, *C* conditional class probabilities *Pr(Class$_i$|Object)* should also be predicted for each grid cell. During test time, class-specific confidence scores for each box are achieved by multiplying the individual box confidence predictions and the conditional class probabilities:

$$Pr(Object)*IOU_{pred}^{truth}*Pr(Class_i|Object) = Pr(Class_i)*IOU_{pred}^{truth}$$

We refer the reader to [8] for the detailed description of the loss function which is optimized during the training.

Knowledge graphs

A knowledge graph *(V,T)* is defined as a union of a set of nodes *V* and a set of directed triples $T \subseteq V \times P \times V$ that is built over a set of predicates *P*. A node represents a certain real world object or concept. A triple $(u,p,v) \in T$ indicates a subject-predicate-object relation. The starting node *u* is the subject, *v* is the object and *p* is the predicate which may contain carries some additional information regarding the nature of the relation between those nodes. Once the detected video objects are mapped to the knowledge graphs, their relationship to each other can be quantified. For the purpose of anomaly detection, we are interested in the closeness of the entities to each other represented in semantic similarity. There have been proposed a number of semantic similarity metrics, which can be broadly divided into corpus-based or knowledge-based approaches [31].

Specific type of neural network architecture and knowledge graph structure can influence the final output of the model. The contents of knowledge graph and the amount of entities which are encoded will impact the resulting similarity scores between the objects. For general outside surveillance video footage as used in this paper, it is most reasonable to use the commonsense knowledge graph. However, for some specific environment such as an agricultural setting or a factory it would be practical to construct a knowledge graph that encodes the entities which belong to the specific domain of interest.

IV. Experiments and Results

The proposed framework for image or frame-by-frame video analysis is built with a pre-trained neural network for object detection and recognition.

Dataset description. The conducted experiment is performed with the Tbilisi 2015 flood video dataset. The video data contains recording from the surveillance cameras of the city streets affected by a large flood along with the wild animals that escaped from the zoo during the flood. The pre-trained solutions that had not been specifically trained for detecting wild animals in the city scenes are not able to recognize them as an anomaly. We perform object detection with the YOLO neural network and identify 22 different classes of objects throughout all the available video. For each frame, we compute semantical similarity between every pair of objects in the frame, whether these are traditional city objects ('car', 'truck', 'traffic light', 'person') or objects which should be classified as an anomaly ('bear', 'hippopotamus') to discover that the wild animals consistently produce significantly lower similarity scores.

Framework and pipeline. For the framework pipeline, we have used the following off-the-shelf components:

1. Pre-trained neural network for object recognition. We used YOLO v4 to identify 22 different classes with an average of 5.2 objects per frame which belonged to 3.7 different classes per frame on average.

2. For the knowledge graph framework for computing distance between the identified entities we have used ConceptNet, a freely available semantic network [27] that contains common entities from real life scenarios, such as objects in the streets, animals, etc. ConceptNet encodes over 13 million links between concepts and includes multilingual representation.

For each frame, we detect the available objects with neural network. Additional entities that are related to the video can be obtained and included from the metadata (geolocations for the video, timestamps). For each pair in the set of the detected objects of a specific frame, their similarity is calculated using the ConceptNet knowledge graph. If a certain object consistently gives a lower similarity score for each pair, it is labeled as a potential outlier.

The proposed solution can be used in unsupervised settings and does not need a training anomaly set to be operational. For testing our implementation, however, we have manually labeled the frames that contained escaped wild animals in the city, as anomalies. The method detected all the frames that contained anomalous animals in the city scenes even if the animals themselves were not correctly classified by the neural network. For example, in some frames the hippopotamus was incorrectly recognized as an elephant due to the low resolution and the object being located at a distance in the camera footage. However, since both objects are located at an equivalently long distance from the rest of the objects in the analyzed frame in the knowledge graph, these errors do not influence the final set of outliers. This demonstrates that the proposed method is stable and robust in the presence of potential misclassifications. The knowledge graph has the major influence on the method accuracy while the neural network errors can be mitigated by the preserving the ratio in the semantic similarity scores. Thus, in the situation when there is a trade off between computational resources allocated to the neural network or the knowledge graph, the latter should be given a priority. From the results presented in the Figure 4, it can be concluded that reducing the neural network accuracy does not bring down the overall number of correctly detected anomalies.

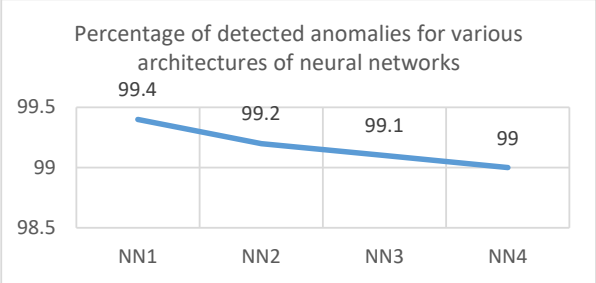

Figure 4. Relationship between accuracy of the pre-trained neural network and percentage of detected anomalies. NN1 represents a neural network with 99.4% mean average precision, NN2 is a neural network with 99.2% mAP, NN3 and NN4 have 99.1% and 99% mAP respectfully

V. Conclusions and Future Work

The framework we presented in this paper consists form the three major modules: deep neural network for object detection and recognition, knowledge graph with the entities from the video, and an anomaly detection module which identifies semantic outliers using the obtained scores. We have demonstrated that the proposed framework can detect the anomalous objects which are semantically different from the rest of the objects in the frame on a frame-by-frame basis. Varying the hyperparameters for the pre-trained neural networks and the threshold for setting aside the most dissimilar object, we were able to reach 99.4% of anomalous object detection. Each classification of the anomalous entity is quantified by corresponding

As a future work, a comprehensive solution can be developed and current framework extended with the following modules:

Additional data modalities module. Data from other modalities such as text and sensor can be included as a source of knowledge. For our dataset, it is possible to extend it with the text transcripts of the news reports. Extracting objects from video and text data simultaneously expands the system into a robust framework that can be used for extracting mission-relevant situational knowledge on demand from streaming multimodal data [28, 29].

Searching the closest entities in the semantic knowledge graphs in the presence of uncertainty. The knowledge graph size directly affects the accuracy of the proposed method. It is imperative that the graph contains the entities detected on the video. The approximation algorithm can be used if the objects detected on the video cannot be located in the knowledge

graph. In this case the closest available entities should be selected. Such entities can be determined with the word2vec and similar methods. Research is needed to develop the most effective approaches for such approximation.

The roadmap for future work is summarized in Figure 5.

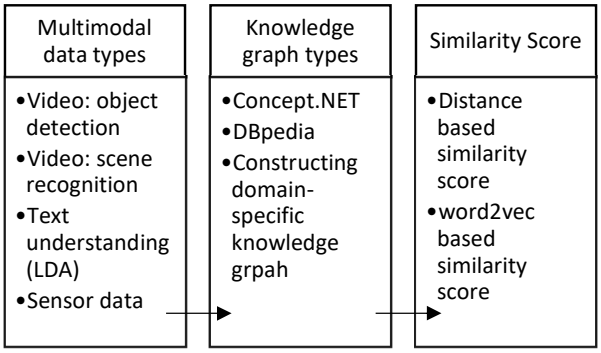

Figure 5. Extended framework with multimodal data sources for anomaly detection with domain specific and domain-agnostic knowledge graphs and semantic networks.

ACKNOWLEDGEMENTS

This research is supported by Northup Grumman Mission Systems University Research Program.

REFERENCES

[1] Karpathy, Andrej, et al. "Large-scale video classification with convolutional neural networks." *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.

[2] Bollacker, Kurt, et al. "Freebase: a collaboratively created graph database for structuring human knowledge." *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. AcM, 2008.

[3] Uyar, Ahmet, and Farouk Musa Aliyu. "Evaluating search features of Google Knowledge Graph and Bing Satori: entity types, list searches and query interfaces." *Online Information Review* 39.2 (2015): 197-213.

[4] Marino, Kenneth, Ruslan Salakhutdinov, and Abhinav Gupta. "The more you know: Using knowledge graphs for image classification." arXiv preprint arXiv:1612.04844 (2016).

[5] Olatunji, Iyiola E., and Chun-Hung Cheng. "Video Analytics for Visual Surveillance and Applications: An Overview and Survey." *Machine Learning Paradigms.* Springer, Cham, 2019. 475-515.

[6] Swann-Quinn, Jesse. "More-than-human government and the Tbilisi zoo flood." *Geoforum* 102 (2019): 167-181.

[7] Wang, Yingxu, et al. "Cognitive informatics: Towards cognitive machine learning and autonomous knowledge manipulation." *International Journal of Cognitive Informatics and Natural Intelligence* (IJCINI) 12.1 (2018): 1-13.

[8] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[9] Seidel, Raimund. "On the all-pairs-shortest-path problem in unweighted undirected graphs." *Journal of computer and system sciences* 51.3 (1995): 400-403.

[10] Ahmed, Faruk, and Aaron Courville. "Detecting semantic anomalies." arXiv preprint arXiv:1908.04388 (2019).

[11] Bertino, Elisa, et al. "Redefining Data Transparency: A Multidimensional Approach." *Computer* 52.1 (2019): 16-26.

[12] Gunning, David. "Explainable artificial intelligence (xai)." *Defense Advanced Research Projects Agency (DARPA), nd Web* (2017).

[13] Bouwmans, Thierry, and El Hadi Zahzah. "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance." *Computer Vision and Image Understanding* 122 (2014): 22-34.

[14] Xu, Dan, et al. "Detecting anomalous events in videos by learning deep representations of appearance and motion." *Computer Vision and Image Understanding* 156 (2017): 117-127.

[15] Xu, Dan, et al. "Learning deep representations of appearance and motion for anomalous event detection." *arXiv preprint arXiv:1510.01553* (2015).

[16] Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2015.

[17] Chong, Yong Shean, and Yong Haur Tay. "Abnormal event detection in videos using spatiotemporal autoencoder." *International Symposium on Neural Networks*. Springer, Cham, 2017.

[18] Xingjian, S. H. I., et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." *Advances in neural information processing systems*. 2015.

[19] Bradley J Rhodes, Neil A Bomberger, Michael Seibert, and Allen M Waxman. Maritime situation monitoring and awareness using learning mechanisms. In Military Communications Conference, MIL- COM, pages 646–652. IEEE, 2005.

[20] Bradley J Rhodes, Neil A Bomberger, and Majid Zandipour. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. In Information Fusion, 2007 10th International Conference on, pages 1–8. IEEE, 2007.

[21] Datta, Ankur, Mubarak Shah, and N. Da Vitoria Lobo. "Person-on-person violence detection in video data." *Object recognition supported by user interaction for service robots*. Vol. 1. IEEE, 2002.

[22] Kooij, Julian FP, et al. "Multi-modal human aggression detection." *Computer Vision and Image Understanding* 144 (2016): 106-120.

[23] Srivastava, Nitish, Elman Mansimov, and Ruslan Salakhutdinov. "Unsupervised learning of video representations using lstms." *arXiv preprint arXiv:1502.04681* (2015).

[24] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems*. 2013.

[25] Chemudugunta, Chaitanya, Padhraic Smyth, and Mark Steyvers. "Modeling general and specific aspects of documents with a probabilistic topic model." *Advances in neural information processing systems*. 2007.

[26] Wei, Xing, and W. Bruce Croft. "LDA-based document models for ad-hoc retrieval." *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2006.

[27] Havasi, Catherine, Robert Speer, and Jason Alonso. "ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge." *Recent advances in natural language processing*. Philadelphia, PA: John Benjamins, 2007.

[28] Palacios, Servio, et al. "WIP-SKOD: A Framework for Situational Knowledge on Demand." *Heterogeneous Data Management, Polystores, and Analytics for Healthcare.* Springer, Cham, 2019. 154-166.

[29] Stonebraker, Michael, et al. "Surveillance Video Querying With A Human-in-the-Loop."

[30] Fang, Yuan, et al. "Object detection meets knowledge graphs." (2017).

[31] Zhu, Ganggao, and Carlos A. Iglesias. "Computing semantic similarity of concepts in knowledge graphs." *IEEE Transactions on Knowledge and Data Engineering 29.1 (2016)*

[32] Jia, Bin, et al. "Pattern Discovery and Anomaly Detection via Knowledge Graph." *2018 21st International Conference on Information Fusion (FUSION). IEEE, 2018.*

[33] Akoglu, Leman, Hanghang Tong, and Danai Koutra. "Graph based anomaly detection and description: a survey." *Data mining and knowledge discovery 29.3 (2015): 626-688.*

[34] Chaudhary, Anshika, Himangi Mittal, and Anuja Arora. "Anomaly Detection Using Graph Neural Networks." *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, 2019.*

[35] Jiang, Jianguo, et al. "Anomaly Detection with Graph Convolutional Networks for Insider Threat and Fraud Detection." *MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM). IEEE, 2019.*

[36] Xiao, Qingsai, et al. "Towards Network Anomaly Detection Using Graph Embedding." *International Conference on Computational Science. Springer, Cham, 2020.*