Severity level assessment from semantically fused video content analysis for physical threat detection in ground segments of space systems *

Gerasimos Antzoulatos[1 [0000−0002−7424−8928]], Georgios Orfanidis[1 [0000−0003−2908−7843]], Panagiotis Giannakeris[1 [0000−0002−3774−3161]], Giorgos Tzanetis[1], Grigorios Kampilis-Stathopoulos[1], Nikolaos Kopalidis[1], Ilias Gialampoukidis[1 [0000−0002−5234−9795]], Stefanos Vrochidis[1 [0000−0002−2505−9178]], and Ioannis Kompatsiaris[1 [0000−0001−6447−9020]]

Information Technologies Institute (ITI) - Centre for Research and Technology Hellas (CERTH), Thermi-Thessaloniki, Greece
{gantzoulatos, g.orfanidis, giannakeris, tzangeor, grigstat, nikokopa, heliasgj, stefanos, ikom} @iti.gr

Abstract. Disaster risks related to natural hazards are evolving gradually, albeit accelerating over time, the human-made and cyber threats are changing rapidly exploiting the increasing progress in technologies and the complex, highly interlinked, modern environment of critical infrastructures. Therefore, as these threats have been intensifying, the actions to strengthen the resilience of critical infrastructures should be step up, by understanding their complex systems as well as the multi-risks nature. In this landscape, the aim of this work focuses on proposing a framework that enables the identification of potential human-made threats, created by the usage of natural means and captured by heterogeneous sources (CCTV, UAV, etc.). Advanced machine learning techniques provide analysis of events and useful information, which are fused semantically and estimate the severity level of the potential attack, serving the needs for real-time monitoring and mitigating the risk.

Keywords: Risk Assessment · Critical Infrastructures · Human-made threats · Video-based Object Detection · Face Detection and Recognition · Knowledge-based Representation · Severity level estimation

1 Introduction

Nowadays, the crisis panorama has changed and diversify increasingly from "traditional" crises generated by natural hazards to technology-driven crises generated by cyber-attacks, or a combination of them ([8], [9]). The unexpectedly large scale of the extreme natural events in terms of their severity and frequency, the trans-boundary and cross-sectoral nature of new or unprecedented crises, compose a challenging and changing landscape in disaster and risk management [2].

* This work has been supported by the EC-funded H2020-883284 7SHIELD project.

In Global Assessment Report on Disaster Risk Reduction 2019, has been underlined the need to move beyond the conventional definition for the disaster risk, re-examine and re-assess the risk, by taking into consideration the pluralistic nature of it: in multiple dimensions, at multiple scales and with multiple impacts [7]. Furthermore, the rise of new technologies, from one side intensifies the potential threats and attacks and from the other, provides empowered solutions to address them and strengthen the resilience in human societies and Critical Infrastructures (CI).

Recent technological innovations like IoT, 5G, unmanned aircraft vehicles, and artificial intelligence have brought immense benefits and contributed further efficiencies to CI operations. However, they have posed serious threats facilitating the malicious actors interested in disrupting CI operations. Particularly, in the CIs which are becoming increasingly complex, automated, and interconnected, thereby new vulnerabilities have been introduced exposing them to malicious physical and cyber-related activities ([8], [9], [13]). Hence, lately the NIS Directive, has been revised in order to extend its scope and include more sectors and services as either essential or important entities (NIS2[1]).

Object detection is considered one of the fundamental fields of computer vision. The detectors can be roughly divided into 2 categories: the two phase ones and the single-phase ones. The former include an extra sub-network which is responsible for proposing bounding boxes. The more prestigious work in the former category is Faster R-CNN [19] while for the latter category are Single Shot Detector (SSD) [17] and You Only Look Once detector (YOLO) [18] (which has actually spawn a family of detectors). The two-phase detectors are considered more robust and effective but also less efficient while the single-phase ones are lighter, more efficient and less effective. Over the years new architectures have emerged which attempt to combine the best of the practices proposed. Such a work is EfficientDet [25] which is based on an efficient backbone, EfficientNet [24] and a bi-directional intra-level feature fusion.

For activity recognition also the focus is on deep learning techniques, since they provide the state-of-the-art performances. One of the first attempts was the Two-stream algorithm [22] which combines two different streams (visual and depth streams in order to increase performance and collect features from both spectra. Also, another monumental work is 3D ResNet [11] which tries to adopt the success of ResNet networks [12] to temporal spectrum by expanding 2D ResNets to the temporal dimension also.

Face recognition depends heavily on deep learning methodologies to achieve significant boost in performance. In this class of algorithms, deep feature extractors are used to generate face representations, tuned for pose and illumination invariance, from the plethora of the available training data rather than from low-level hand-crafted features. Siamese networks for deep metric learning were proposed in the work of [4], which was one of the initial attempts to leverage deep learning. A Siamese network works by extracting features separately from

[1] `https://digital-strategy.ec.europa.eu/en/library/revised-directive-security-network-and-information-systems-nis2`

two modes (inputs), with two identical Convolutional Neural Networks (CNNs), taking the distance between the outputs of the two CNNs as dissimilarity. In a similar fashion with face detection works, facial parts were processed separately in cascade networks, as in the work of [23]. Soon after, the focus shifted heavily towards improving the deep metric learning methodology, which led to significant performance improvements [6]. Experimentation with novel face similarity measures dominates the undertaken effort in these works. Moreover, discriminant face representations are characterized by smaller maximal intra-class distance and minimal inter-class distance in the embedding space, thus, novel CNN loss functions are meticulously explored as well, in order to find the most appropriate for the task.

Although the application of Machine Learning methodologies to tackle specific problem areas in disaster risk management dates back to a recent couple of decades, however, significant challenges still need to be addressed [13]. Machine Learning methods have penetrated in a descriptive and/or predictive manner in all the phases of disaster/crisis management, contributing in various ways to the assessment of the hazard, exposure and vulnerability from natural and human-made disasters [27]. Hence, one of the main challenges concerns the lack of required training data which limits the utilisation of the machine learning algorithms to be trained in order for the latter to be able to predict or assess the risk of a crisis event. Motivating by this gap, the proposed annotation tool aims to involve the experts in the Satellite ground segments domain, by mapping their experience and knowledge into the characterisation of hypothetical extreme physical (natural or human-made) events in terms of their severity and impact.

The continuous growth of semantic web technologies provides several ontology-based approaches in several domains. For this task, the categorisation of the domains includes the events and observations, the crisis management and the cyber-physical threats and vulnerabilities. In particular some representative ontologies for each domain respectively, that influenced the process and the methodological approach for our framework include SSN [5] and SOSA [15] for mapping of sensors and their observations, properties and features of interest; MMF [10] an ontology developed in the context of managing sensor assignment to mission; finally MOAC [16] and SoKNOS [1] with wide field of application in crisis management and response. Our ontological representation is tailored to the protection of ground segments of space systems.

In this work, we focus on the detection and monitoring of physical threats generated by human-made malicious activities on ground segments of space systems. The potential attacks are classified and assessed in terms of their severity level and potential consequences in the ground segments, supporting in this way the decision-making processes to mitigate the risks. Machine learning advances are the core aspect of our approach as innovative deep-learning methodologies analyse multimedia content from videos, aiming to detect malicious objects and suspicious activities of identified and potentially unauthorised persons in restricted areas. Finally, the semantic fusion of information leads to the real-time monitoring and assessment of the potential attack's severity level are carried out

by utilising machine learning methods. Due to the lack of adequate annotated datasets in automatic risk assessment supervised methods, we propose an annotation tool that aims to engage the community of users and experts in the domain of the protection of ground segments of space systems.

2 Methodological Framework for physical attack detection and response

In this work, our intention is to highlight some aspects of the above framework, especially those that detect physical attacks, fuse semantically the identified malicious events, and assess the severity level of those attacks. The proposed framework combines tools (*Detection Layer*) for detection and recognition of objects, faces, and activities from video-based content, that obtained from surveillance systems (CCTVs cameras) or cameras on the UAVs. After the detection, the generated alerts of the events are combined, homogenised and semantically indexed in the Knowledge Base (*Fusion Layer*). The enriched information is propagated to the Crisis Classification module which is responsible to estimate the severity level of the event and propagate the results to the CI operators (*Decision Layer*) to support decision-making and mitigation actions for timely response to the physical threat. In Fig. 1 the workflow of information as well as the interactions between modules in various levels are illustrated. In the following subsections a more detailed description of the functionalities of each module is exhibited.

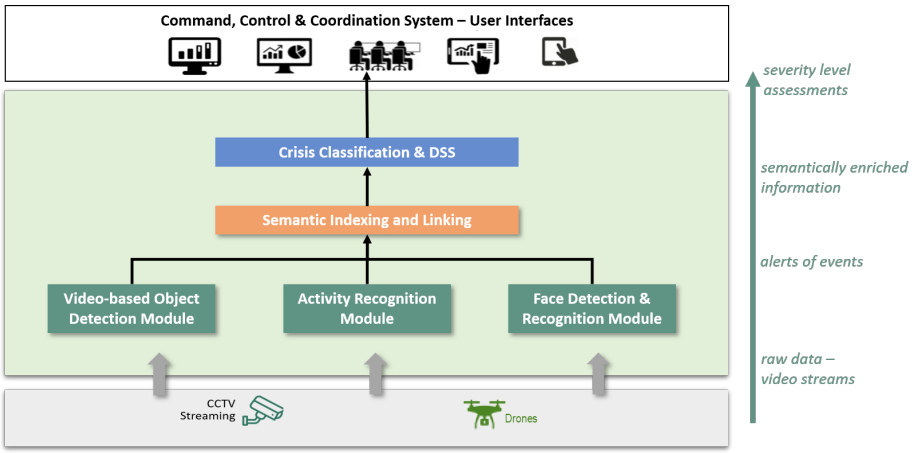

Fig. 1. The proposed decision support framework

2.1 Video-based Object Detection and Activity Recognition

The surveillance of ground segments of space systems is a vital issue for their secure and seamless operation since new threats seem to arise and some of them especially focusing on those infrastructures. A huge asset to the latter is the visual understanding of the surrounding area. The Video-based Object Detection (VOD) and Activity Recognition (AR) modules are efforts to aid towards this aim.

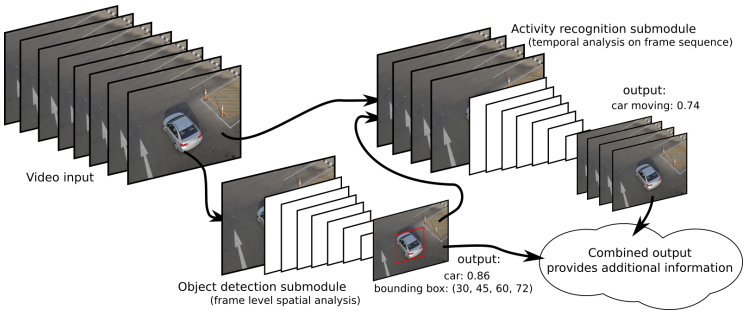

Fig. 2. High-level view of combined VOD-AR architecture

First, the VOD module utilise deep learning techniques in order to visually locate and identify the objects of interest inside the ground segment of space systems. The input of the module are video streams which are being processed by large in width but mainly in depth networks. The analysis provide the system with an initial interpretation of the monitored area regarding the objects appearing in it. Although, the initial analysis is performed on a frame level, an interconnection with consecutive frames is also provided, augmenting the capabilities of the system to clearly isolate true threats from false positive ones. The actual outcome of the VOD module is a group of bounding boxes around each detected object of interest accompanied by a confidence score, which reveals how certain is the network for this detection, and label to denote the class the object belongs to. Since each detection is performed on a specific frame a spatio-temporal association of the detected objects across consecutive frames can be deducted, and, this correlation can be further feed to the relevant AR module. AR module is responsible for identifying an activity given a specific frame span (or equivalently a time span and a video from where the temporal boundaries can be deducted). Thus, VOD can function as a trigger for AR module if certain conditions are met. Such conditions could be a combination of objects being detected, such as a person and an object like a bag, a vehicle to specific location etc. Of course, in a more generic mode, all detected objects involving in potential activities could be forwarded to the AR to decide the existence of any potential harmful and suspicious activity. Summarising, the output of the Video-based Object detection is:

– Awareness of surroundings via detected objects and individuals
– Bounding boxes and class label for each instance of interest

while for Activity Recognition, the output is:

– Awareness of surroundings via recognized activities
– Label for each activity along with the participating objects

The innovative part of the VOD module is the combined object detection and activity recognition output. Since, object detector by default do not involve any temporal information and the activity recognition do not consider any spatial one, their combination can produce a more thorough analysis of the surroundings which could include additional information. The idea of combining the two submodules is decipted in Figure 2 where each submodule produces its own outputs but their is also interconnection between them as VOD feeds AR submodule.

2.2 Face Detection and Recognition

In ground segments of space systems, it is common to restrict access on certain areas to unauthorised personnel. Typically, access is granted manually by security guards, or with electronic access control systems via identity cards. However, these control mechanisms may be vulnerable to identity fraud attacks. For example, someone could get access to a building or an area by using a lost or stolen card. Therefore, the traditional solutions, when used alone, cannot guarantee maximum security. Our solution is designed to assist in access control systems using automatic facial recognition.

The Face Detection and Recognition (FDR) module ensures that restricted access to facilities is under secure control. At the same time, this module may also assist in intrusion detection systems, by notifying about unauthorised access to areas of interest. Within this objective, Satellite Ground Segments and general Critical Infrastructure are protected from hazardous activities of unauthorised trespassers while the corresponding personnel and their daily activities are also secured.

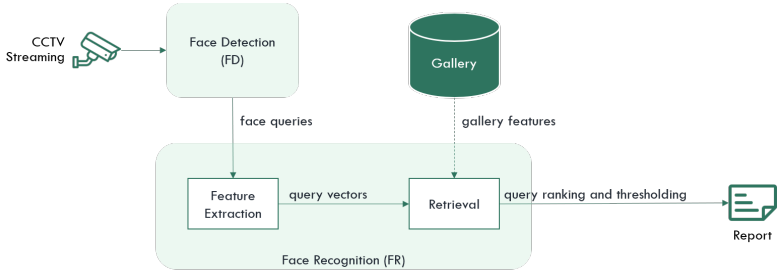

Fig. 3. High-level FDR architecture

The overall approach of FDR is naturally split into its two cooperating tasks and is shown in Fig. 3. The module is initialized with a video stream and it is designed to process single video frames in a serial processing pipeline (one after another). Processing begins on the Face Detection (FD) component. FD is responsible to detect patches inside the input frame where faces are tightly enclosed. The acquired face patches are instantly characterized as unknown and are immediately provided to the Face Recognition (FR) component for further processing. FR takes additional input from a pre-existing gallery of known faces and tries to match the detected faces with the ones from the gallery. The gallery images belong to authorized personnel with unrestricted access in the area covered by the CCTV camera. After the recognition process, detailed reports can be produced with the detection and recognition metadata, e.g., alarm notifications of potential unauthorised access, list of recognised identities with attached timestamps for monitoring access to critical assets, enhanced video data with bounding boxes showing the detected faces for visualisation in command and control dashboards, etc.

2.3 Semantic Indexing and Linking

The Knowledge Base (KB), is a knowledge representation model for semantically representing concepts relevant to the cyber-physical threats. The goal of the KB framework is to research and develop technologies for semantic content and sensor input modelling, integration, reasoning and question answering.

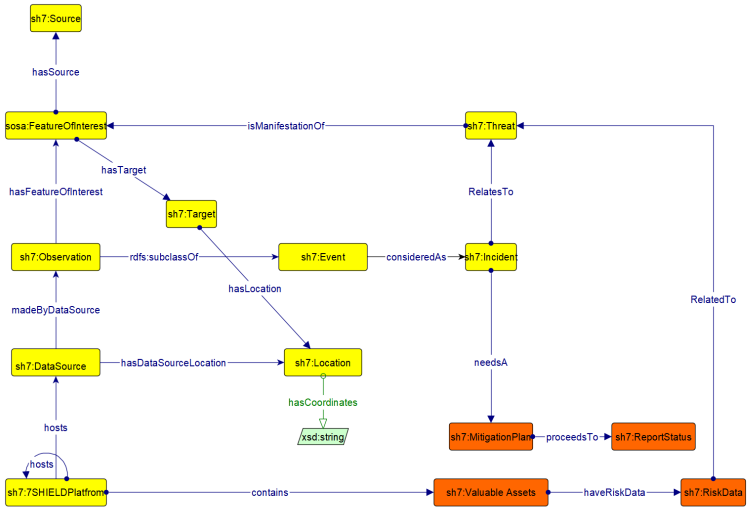

Fig. 4. High Level overview of 7SHIELD ontology

The models that are created will constitute for the reasoning mechanisms taking into account the ontology vocabulary and infrastructure for capturing and storing information related to the 7SHIELD[2] application domain, such as: (a) Observation and Events (e.g. data collection from face recognition/detection, multimodal automated surveillance, drone detection), (b) C/P security (e.g. cyber detection, correlation services output), (c) Mitigation and response plans (e.g. First responder teams, UAV neutralisation). The 7SHIELD Knowledge Base (KB)", can be also called 7SHIELD ontology modelling, will be described below. The 7SHIELD ontology was based on Semantic Sensor Network (SSN) ontology and the OWL language was used.

In Fig. 5 the 7SHIELD ontology, that consists of classes in high level and their entities, is illustrated using Protégé[3], which is an open-source ontology editor and framework for building intelligent systems.

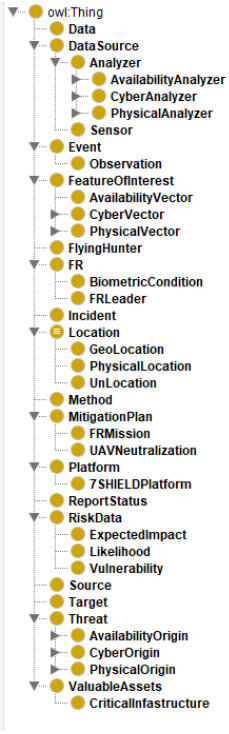

Fig. 5. List of classes as they are viewed in Protégé

[2] https://www.7shield.eu
[3] https://protege.stanford.edu/

- **DataSource**: This class represents data that have been analyzed and a result has been extracted
- **Event**: This class represents one of the primaries of the overall data model of the information sharing environment. Event is an abstract entity which has a subclass, the Observation
- **Location**: This class represents the place or position that something is in or where something happens. The class is further divided into 3 subclasses (PhysicalLocation, GeoLocation, Unlocation).
- **Target**: This class represents an object of attention or attack.
- **7SHIELD Platform**: This class hosts other entities, particularly Sensors, Detectors, Samplers
- **ReportStatus**: Its purpose is to make a report when triggered from an event.

Finally, the purpose of the data converter module is to receive JSON data as input and accordingly form the TURTLE Resource Description Framework (RDF) data as output, for mapping them the RDF triplestore. TURTLE[4] is a syntax and file format for expressing data in the RDF data model. The JSON data should be in the appropriate format in order to be converted to semantic data (RDF triplets).

2.4 Crisis Classification & DSS Module

The main goal of the Crisis Classification (CRCL) & DSS module (Fig. 6) is to enhance the decision-making processes, by providing real-time assessments of the severity level of an ongoing physical and/or cyber-attack in critical satellite and ground segments. To achieve this goal, a multi-level fusion approach is developed which encompass methodologies for Information and Decision fusion.

At the *Information Fusion* level, the real-time (or "near" real-time) information, generated by the fusion of heterogeneous data from detection modules, is analysed by utilised machine learning techniques that are able to estimate the severity level of a malicious event. Then, at the *Decision Fusion* level, decision-making approaches will be tailored aiming to enrich the outcomes of the Information Fusion level semantically with information extracted from Knowledge Base. Hence, this process will estimate accurately, interpret and provide assessments in terms of the severity level and classify the crisis events generated by C/P attacks. This approach and the CRCL module are easily adjustable to fuse information from various available modules depending on the field of application.

3 Experimental validation and evaluation

3.1 Evaluation of the Detection layer

Visual Object Detection for Activity Recognition. We have been experimenting with various object detection models in order to achieve a working

[4] `https://www.w3.org/TR/turtle/`

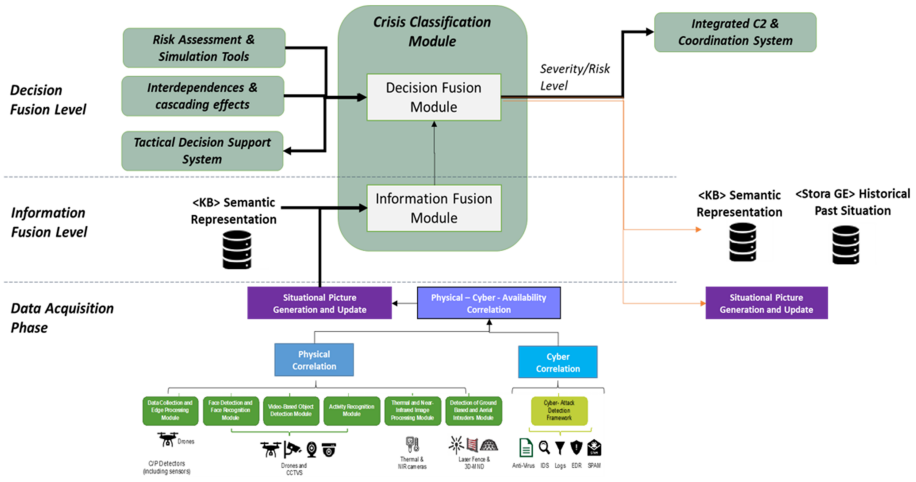

Fig. 6. CRCL module in 7SHIELD framework

solution. First, a Faster R-CNN two-phase detector model has been trained on a specially collected dataset of over 20k samples. This model can detect 6 classes: a UAV class, a Car class, a Bus class, a Truck class (which also include Van instances), a combined Motorcycle/Bicycle class, and a Person class. We have also experimented with a lighter (but less effective) model, namely an EfficientDet ($\phi = 0$) model, which also included 6 classes (but with a few differences): a Car class, a Bus class, a Truck class, a Bicycle class, a Motorcycle class, and a Person class and was trained on 10k samples. The evaluation was performed on a distinct specially chosen dataset of roughly 200 samples in order to cover the requirement being set.

Table 1. VOD results using 2 different architectures

Object detection results using Average Precision (AP)		
Faster RCNN		
0.75330 (UAV)	0.57315 (bus)	0.75726 (car)
0.73409 (moto-bike)	0.82152 (person)	0.53351 (truck)
mean AP: 0.6954		
EfficientDet $\phi = 0$		
0.4563 (person)	0.4668 (car)	0.3438 (bicycle)
0.5562 (bus)	0.3968 (motorcycle)	0.3790 (truck)
mean AP: 0.4332		

As a first note for the results in Table 1, the results are not completely comparable since they include somehow different classes. Nevertheless, it is clear that the two-fold detector (Faster R-CNN) seems to perform better in the core

detection part. The reported Average Precision values are much higher than its counterpart EfficientDet. On the other hand regarding the efficiency of the model EfficientDet is much lighter and faster by one order of magnitude. Regarding the results for AR submodule it is not so easy to be evaluated because they are highly dependent to the output of the VOD submodule.

Face Detection and Recognition. The experiments in this section were conducted with the aim to (a) deploy deep learning face detection and recognition models as a means of testing the development platform, (b) replicate and confirm the published evaluation results on public benchmarks and (c) make performance comparisons and draw conclusions about the state-of-the-art. For each task, three approaches were selected to represent the current state-of-the-art landscape, i.e., for face detection, (i) TinyFaces [14], (ii) PyramidBox [26], (iii) DSFD [28] and for face recognition, (i) Facenet [20], (ii) PFE [21], (iii) Arcface [6]. Each one was evaluated in a benchmark dataset, appropriate for the task. Specifically, for face detection the WIDER FACE benchmark was selected, and for face recognition the LFW.

WIDER FACE [29] is a face detection benchmark dataset. It contains over 30000 images which mostly show people participating in various activities of everyday life based on 61 event classes. The human faces appear with a high degree of variability in scale, pose and occlusion. For each event class, predefined splits consisting of 40%/10%/50% of the total amount of data exist as training, validation and testing sets respectively.

Table 2. Face detection state-of-the-art evaluation

Method	WIDER FACE AP (%)
TinyFaces (2017) [14]	90.7
PyramidBox (2018) [26]	94.3
DSFD (2019) [28]	95.5

The Labelled Faces in the Wild (LFW) [3] dataset is a database of face photographs designed for studying the problem of unconstrained face recognition. The data set contains more than 13,000 images of faces collected from the web. The people that appear in this dataset are known public figures like politicians, athletes, actors, musicians and other various celebrities.

Table 3. Face recognition state-of-the-art evaluation

Method	LFW Accuracy (%)
Facenet (2015) [20]	99.4
PDE (2019) [21]	99.6
Arcface (2019) [6]	99.7

The evaluation metric for face detection is Average Precision. It is taken by calculating the area under the Precision-Recall curve. Precision is defined as the proportion of true positives (TP) out of all the detected faces and Recall as the proposition of true positives out of all the annotated faces. In other words, precision measures the accuracy of the detector and recall measures its ability to retrieve the existing faces. Whether a bounding box detection counts as a TP is decided based on its overlap with a ground truth box. The overlap is measured by the Intersection over Union (IoU) threshold. Thus, detected faces must have a good alignment with true faces in order to be considered correct. The Precision and Recall metrics are calculated for every alignment threshold (from most relaxed to most strict) to draw the Precision-Recall curve. The evaluation metric for face recognition is Accuracy. The dataset is split to 10 equal parts, where the first 9 are used for cross validation in order to select the optimal distance threshold to achieve top accuracy. The 10th part is the test set from which pairs of queries are given and the model decides if they belong to the same person or not. This is a standard strategy for evaluating face verification models, but it is also a good indicator of face recognition performance as well.

Tables 2 and 3 show the performance comparison of face detection and recognition SoA methods respectively. There is an overall good agreement with published results, with maximum deviation at 0.3%. Regarding face detection performance in WIDER FACE, the three methods achieve high average precision, especially the more recent approaches. From the ones that focus on leveraging surrounding face context, the PyramidBox is the most superior. Regarding face recognition performance in LFW, all methods perform extremely well, which may indicate both superior performance of SoA and dataset saturation.

3.2 Validation of the Fusion layer

We also present the metrics about the current version of the 7SHIELD ontology, we used the OntoMetrics tool, an online framework that evaluates the ontology based on predefined metrics. The following tables present the results of the aforementioned process. The Figure 7 contains the base metrics which show the quantity of the ontology; numbers of triples, classes, object and datatype properties and individuals.

3.3 Annotation Tool for the validation of the Decision layer

As mentioned above, the main issue in the utilisation of Machine Learning techniques is the lack of annotated datasets, namely datasets that assess the severity level of an attack with the specific characteristics of the attack (physical or cyber). To overcome this, we designed and developed the Annotation tool that aims to capture the knowledge and experience of experts in a qualitative, simple, fast and user-friendly way. The main idea of this tool is to generate scenarios of physical or cyber attacks in specific locations/assets in pilot sites and request experts to characterize those scenarios in terms of likelihood of the attack and

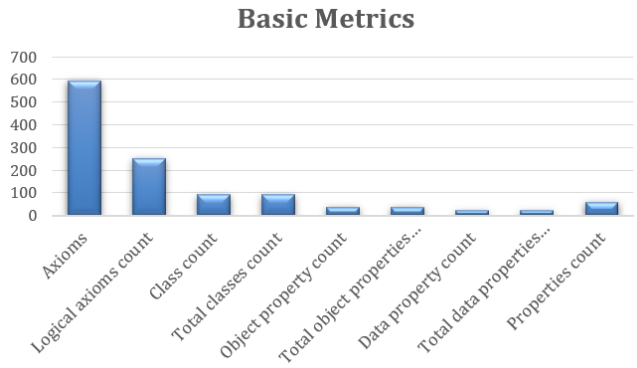

Fig. 7. Basic ontology Metrics

potential consequence of it. Combining these two concepts we can assess the severity level.

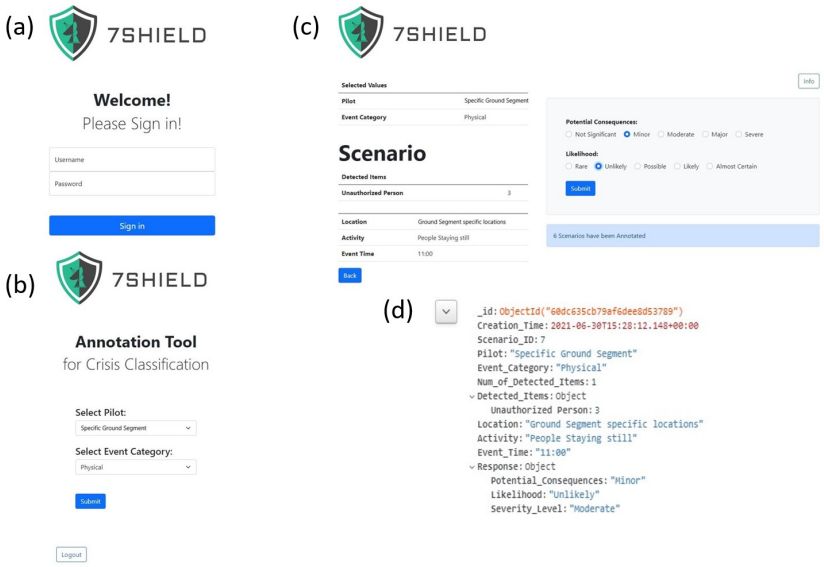

Fig. 8. Annotation Tool; (a)Login page, (b)Selection page, (c)Main page, (d)MongoDB

Annotation Tool is a Web Application. The users can access it through a web browser with an active network connection. The users must first login, using the credentials that were given to them. Then, at the selection screen, the specific Satellite Ground Segment and the Event Category (cyber or physical)

can be selected. Based on this selection random scenarios are generated. The hypothetical scenario is represented under the "Scenario" tag. The users, after studying the random parameters, must select a "Potential Consequence" and a "Likelihood" value. Then, the annotated scenario can be submitted and stored online in the MongoDB database. Automatically, the process continues and the next non-annotated scenario appears. Finally, the estimation of the "Severity level" is carried out, by relying on the risk matrix (Figure 9), that adjusts to the project's needs.

Severity		Potential Consequences				
		Not Significant	Minor	Moderate	Major	Severe
Likelihood	Almost Certain	Moderate	High	Extreme	Extreme	Extreme
	Likely	Moderate	High	High	Extreme	Extreme
	Possible	Low	Moderate	High	High	Extreme
	Unlikely	Low	Moderate	Moderate	High	High
	Rare	Low	Low	Low	Moderate	Moderate

Fig. 9. Risk Matrix used to calculate Severity Level

4 Conclusions

In this work we present an overall framework for the detection, semantic indexing and severity level estimation during physical attack scenarios in ground segments of space systems. Our set of modules includes not only visual analysis technologies but ontological representation and semantic indexing, coupled with a crisis classification module that estimates the level of severity during a physical threat. Finally, the annotation tool which has been developed is planned to be distributed to operators of ground segments of space systems for the creation of ground truth data that will be used in training, validating and testing the future crisis classification algorithms. The annotation tool will also be extended to cyber/physical threats in other critical infrastructures beyond the considered ground segments of space systems.

References

1. Babitski, G., Bergweiler, S., Grebner, O., Oberle, D., Paulheim, H., Probst, F.: Soknos – using semantic technologies in disaster management software. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) The Semanic Web: Research and Applications. pp. 183–197. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)

2. Baubion, C.: Strategic crisis management. OECD Risk Management **23** (2013). https://doi.org/https://doi.org/10.1787/5k41rbd1lzr7-en
3. Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N.: Localizing parts of faces using a consensus of exemplars. IEEE transactions on pattern analysis and machine intelligence **35**(12), 2930–2940 (2013)
4. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 1, pp. 539–546. IEEE (2005)
5. Compton, M., Barnaghi, P., Bermudez, L., GarcÃa-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K.: The ssn ontology of the w3c semantic sensor network incubator group. Journal of Web Semantics **17**, 25–32 (2012). https://doi.org/https://doi.org/10.1016/j.websem.2012.05.003, https://www.sciencedirect.com/science/article/pii/S1570826812000571
6. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4690–4699 (2019)
7. for Disaster Reduction, U.N.I.S.: Global Assessment Report on Disaster Risk Reduction 2019. United Nations (2019). https://doi.org/10.18356/f4ae4888-en, https://www.un-ilibrary.org/content/books/9789210041805
8. EC: Overview of natural and man-made disaster risks the european union may face. https://ec.europa.eu/echo/sites/default/files/overview_of_natural_and_man-made_disaster_risks_the_european_union_may_face.pdf (2020)
9. EY: Evaluation study of council directive 2008/114 on the identification and designation of european critical infrastructures and the assessment of the need to improve their protection. https://op.europa.eu/en/publication-detail/-/publication/118dcd3d-b041-11ea-bb7a-01aa75ed71a1 (2019). https://doi.org/10.2837/864404
10. Gomez, M., Preece, A., Johnson, M.P., de Mel, G., Vasconcelos, W., Gibson, C., Bar-Noy, A., Borowiecki, K., La Porta, T., Pizzocaro, D., Rowaihy, H., Pearson, G., Pham, T.: An ontology-centric approach to sensor-mission assignment. In: Gangemi, A., Euzenat, J. (eds.) Knowledge Engineering: Practice and Patterns. pp. 347–363. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
11. Hara, K., Kataoka, H., Satoh, Y.: Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6546–6555 (2018). https://doi.org/10.1109/CVPR.2018.00685
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
13. Hegde, J., Rokseth, B.: Applications of machine learning methods for engineering risk assessment – a review. Safety Science **122**, 104492 (2020). https://doi.org/https://doi.org/10.1016/j.ssci.2019.09.015
14. Hu, P., Ramanan, D.: Finding tiny faces. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 951–959 (2017)
15. Janowicz, K., Haller, A., Cox, S.J., Le Phuoc, D., LefranÃ§ois, M.: Sosa: A lightweight ontology for sensors, observations, sam-

ples, and actuators. Journal of Web Semantics **56**, 1–10 (2019). https://doi.org/https://doi.org/10.1016/j.websem.2018.06.003
16. Limbu, M., Wang, D., Kauppinen, T., Ortmann, J.: Management of a crisis (moac) vocabulary specification. Web: http://observedchange. com/moac/ns/(zuletzt besucht am: 29-07-2014) (2012)
17. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.: Ssd: Single shot multibox detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science. vol. 9905, pp. 21–37 (9 2016). https://doi.org/10.1007978-3-319-46448-0_2
18. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788 (2016). https://doi.org/10.1109/CVPR.2016.91
19. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**, 1137–1149 (2015)
20. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)
21. Shi, Y., Jain, A.K.: Probabilistic face embeddings. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6902–6911 (2019)
22. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. CoRR **abs/1406.2199** (2014), `http://arxiv.org/abs/1406.2199`
23. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2892–2900 (2015)
24. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (09–15 Jun 2019)
25. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10778–10787 (2020). https://doi.org/10.1109/CVPR42600.2020.01079
26. Tang, X., Du, D.K., He, Z., Liu, J.: Pyramidbox: A context-assisted single shot face detector. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 797–813 (2018)
27. Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., Mestav Sarica, G., Ruangpan, L., Molinario, G., Lallemant, D.: Invited perspectives: How machine learning will change flood risk and impact assessment. Natural Hazards and Earth System Sciences **20**(4), 1149–1161 (2020). https://doi.org/10.5194/nhess-20-1149-2020
28. Wang, C.Y., Liao, H.Y.M., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H.: Cspnet: A new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 390–391 (2020)
29. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: A face detection benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5525–5533 (2016)