A vision on Hybrid AI for military applications

Judith Dijk*[a], Klamer Schutte[a], Serena Oggero[a]
[a]Dept. of Intelligent Imaging, part of TNO - Dutch Organisation for Applied Scientific Research, Oude Waalsdorperweg 63, 2597 AK, The Hague, the Netherlands

ABSTRACT

Application of different Artificial Intelligence technologies is increasing over the past couple of years. At a high conceptual level, we can divide these technologies in two different categories: symbolic and sub-symbolic. The term "Hybrid AI" denotes the combination of symbolic and sub-symbolic AI. By combining both semantic reasoning and data-driven machine learning both human specified and data derived knowledge can be combined in one system.

In this paper we explore the concept of Hybrid AI by the hand of architectural patterns from literature. The added value of the architectural patterns is that they provide a way to discuss the different elements in the processing pipeline. They stimulate discussion what the input and output of the different processing blocks are, and how they work together. When applying the available design patterns to real military imaging applications, we noticed that we needed more detail in the different blocks to specify the type of data or algorithms that are applied. In future work we will investigate how components such as online learning can be presented in this design pattern framework.

We identified the need to further develop this approach with a more intertwined interaction between the reasoning and the data-driven part of the pipelines, and use more world knowledge, domain knowledge and relations between objects in the reasoning part. Improvements are also needed for online learning, where the knowledge of the system performance will be used to ask the users relevant information.

Keywords: Hybrid Artificial Intelligence, architectural patterns, imaging, information extraction

1. INTRODUCTION

Application of different Artificial Intelligence technologies is increasing over the past couple of years. At a high conceptual level, we can divide these technologies in two different categories: Symbolic and Sub-symbolic. Symbolic AI is defined as a knowledge or model-driven form of AI. Knowledge that can be used is 1) knowledge of the world, such as context and external environment, 2) knowledge of the system itself, such as the sensing instruments including setup and calibration, the processing applied and the quality of the results, and 3) knowledge of the application domain, such as casual relations, physical laws and language rules. The main challenge for symbolic AI is how to come up with a knowledge specification which is complete, and valid for a sufficient large set of boundary cases. Sub-symbolic AI, on the other hand, can be defined as data-driven AI. Here the setup is to learn from data and often also supplied labels. A main drawback for sub-symbolic AI is the lack of comprehension of the results. The term "Hybrid AI", as used in this paper, is defined as the combination of symbolic and sub-symbolic AI. By combining both semantic reasoning and data-driven machine learning both human specified and data derived knowledge can be combined in one system.

* Judith.dijk@tno.nl

Dijk, Judith, Klamer Schutte, and Serena Oggero. "A vision on hybrid AI for military applications." Proceedings SPIE, Artificial Intelligence and Machine Learning in Defense Applications. Vol. 11169. International Society for Optics and Photonics, 2019.
http://dx.doi.org/10.1117/12.2551893

In this paper we explore the concept of Hybrid AI by the hand of architectural patterns defined by van Harmelen [1][2]. These architectural patterns are described in section 2. We apply the design patterns on four examples in military imaging applications. These examples are provided in Section 3, together with results of the application. In Section 4 we will draw some conclusion and discussion about this approach of Hybrid AI and the added value of the architectural patterns.

2. ARCHITECTURAL PATTERNS

Van Harmelen depicts different architectural design patterns for AI systems using two different elements: ovals for algorithms and boxes for their input and output [1][2]. The oval algorithms can be SR for symbolic reasoning and ML for data-driven machine learning. The input and output in his scheme are symbolic relational structures (called `sym`) or other data (`data`). The architectural patterns for classical symbolic reasoning systems and classical machine learning systems as defined by van Harmelen are given in Figure 1 and Figure 2.

Figure 1 Classical symbolic reasoning system

Figure 2: Classical machine learning system

A first example of a Hybrid AI pattern is learning with domain knowledge as prior, as presented in Figure 3. An example of this pattern is a Logical Tensor Network [3], where the prior symbolic knowledge can be used to train networks with fewer training data obtaining more robustness against noise.

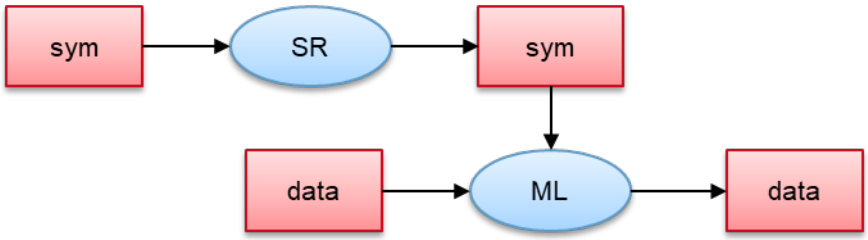

Figure 3 Design pattern for learning with domain knowledge as prior

A second Hybrid pattern is ontology learning [4], where a symbolic structure is learned from data. This ontology is then used in the next step for reasoning. This pattern is shown in Figure 4. This can for example be used for ontology learning from text.

Figure 4 Design pattern for ontology learning

In this paper we adopt this notation and map a number of military imaging examples to these architectural patterns.

3. MILITARY IMAGING EXAMPLES

We have applied these architectural patterns to four military imaging applications in which AI is applied:

- Tank detection, combining a knowledge driven simulation feeding a data driven machine learning application;
- Ship detection, combining transfer learning and engineered feature detectors;
- Ship classification, combining data augmentation, machine learning and generative object models;

- Event detection, combining machine learning for low level tasks with high level event detection.

3.1 Tank detection

A common military task is the detection of a target. In this application the task was to train a deep learning network to detect tanks in drone imagery. However, there is not enough representative training and test data to train this neural net. Therefore, we generated training data using gaming simulation software: Grand Theft Auto 5 [5]. The architectural pattern for this is given in Figure 5. Knowledge is used in the Simulation step to create test and train data, which can be used in the Machine Learning step to generate a deep neural net that is able to detect tanks. The validation is done with real data.

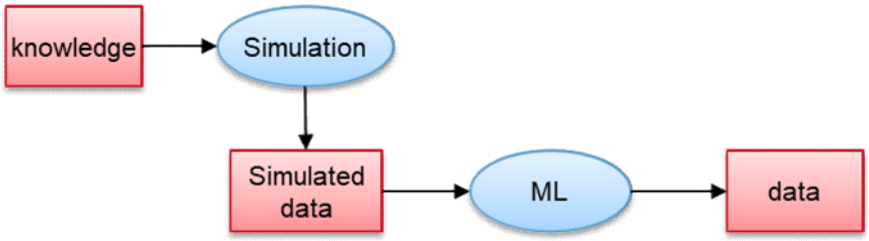

Figure 5: Design pattern for tank detection using simulated data.

An example of the training data is given in Figure 6. Detected tanks in real imagery are shown in Figure 7. The validation showed that the general detection of the tanks was pretty good, although not perfect. In general, these results show that using simulated data for machine learning is a valid approach.

Figure 6: Example of simulated training data for tank detection.

Figure 7: Results of tank detection on real data (retrieved from AnnaNews [6])

3.3 Ship detection

Another target class of interest are ships. Again, training data is scarce, because there are few examples in the infrared bands often used. For ships there are pretrained networks already available, such as a Single Shot Detector (SSD) [7] as

used later, but these typically are trained on data from visual cameras only. Our solution is to use data from the visual domain to learn a detection model, and then transfer this model to the IR domain by fine tuning on a small number of IR examples. The design pattern related to this approach is presented in Figure 8.

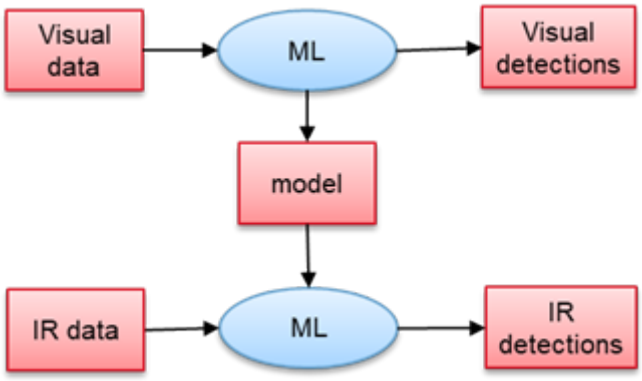

Figure 8 Design pattern for ship detection using transfer learning

We implemented this design pattern [8] using a SSD [7], a convolutional network that produces a set of bounding boxes with associated object class scores. We implemented the SSD using an open source Keras-implementation [9] with a VGG-16 base network and pre-trained on the VOC2007 and VOC2012 trainval datasets [10]. We retrained the network to detect one class ('ship') and trained for 50 epochs with a batch size of four. A leave-one-out procedure was adopted in which for each evaluation dataset an SSD network was trained using all other available data (all development datasets and the evaluation datasets of other scenes) both for MWIR and LWIR. The first three layers of the VGG-16 network were frozen and data augmentation was used to increase robustness against object variability. The data augmentation is in line with the original SSD paper and comprises horizontal flipping, cropping and varying saturation, lighting, contrast and brightness values.

Figure 9 Example for ship detection. Left the annotated data, right the SSD detections. This neural net was first trained on visual imagery, and then retrained with IR data.

For ship detection we see that some ships are better found using deep learning, and other ships better using engineered feature models such as contrast based models. Therefore, we combined the results of the transfer learning approach with this contrast approach, which is able to detect smaller objects, but has an overfit of the detection in case of wake or other structure close to the ship. This design pattern is shown in Figure 10.

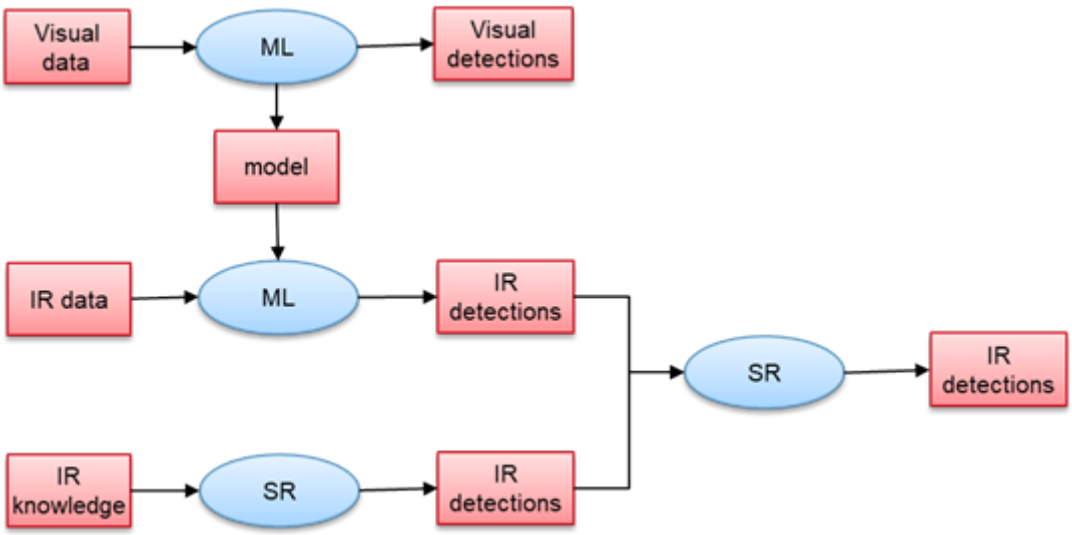

Figure 10: Design pattern for combination of deep learning and semantic reasoning for ship detection

The resulting detections are shown in Figure 11. It can be seen that in this figure both the small and the large ships are detected.

Figure 11: Example of ship detection using a combination of detection method. Left the annotated data, right the detected ships. Note that the large ship at the right is in the part of the image where no detections are done.

3.4 Classification of ships

A third example is the classification of ships. In our approach the final classification is a model-based approach where the contour of a ship is matched to a database. However, the segmentation of the ship and the background, needed to obtain the ship, is done using a learning based neural net. The challenge for such an application is that the amount of available training data is again limited, and that there is a misfit between the training data and the operational data. To be able to train the system properly we augment the available training data [11]. The design pattern for this case is presented in Figure 12. Here an existing segmentation model is retrained using available data and augmented data to obtain a better performing segmentation network. This segmentation is used to retrieve a ship silhouette, which is matched against a 3D database to obtain a ship classification. The 3D database is generated using a model-based generative method based on a silhouette in a source image, typical features of a generic ship and user knowledge about specific ships.

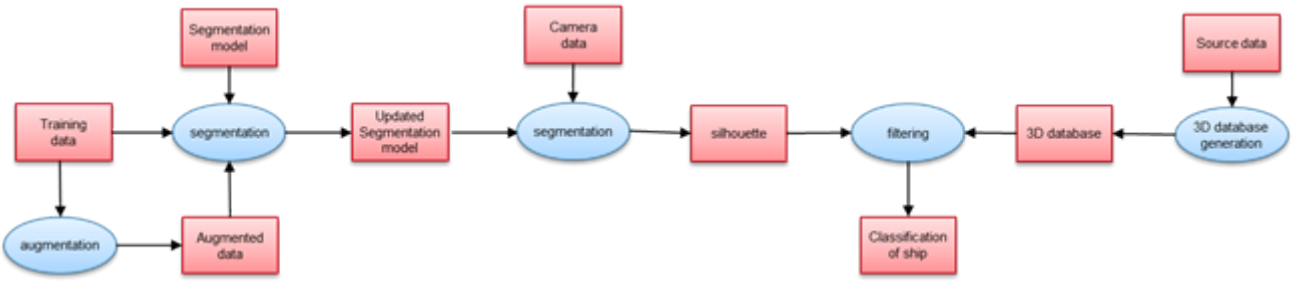

Figure 12: Design pattern for ship classification using data augmentation and 3D database generation

An overview of the ship classification approach is presented in Figure 13 [12]. Based on the camera image a contour is selected which is matched to a database. This provides the best match or a number of best matches, as shown in Figure 13.

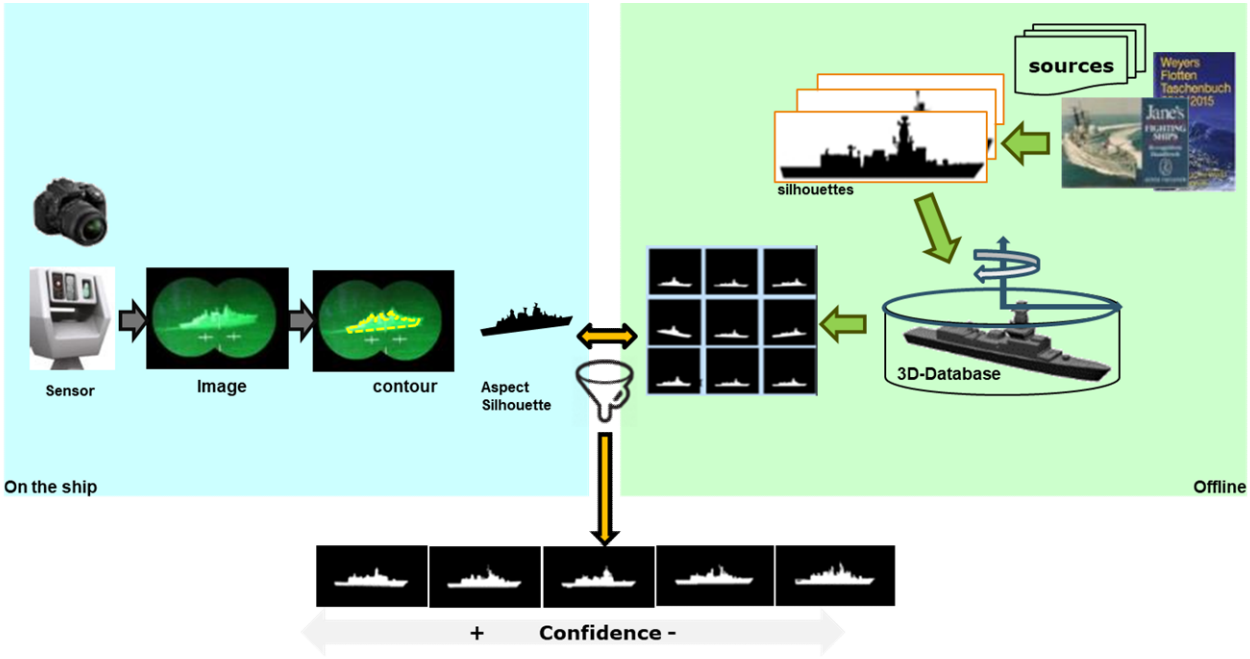

Figure 13 Ship classification: from camera to contour matching. On the left the online detection of a ship using a camera is shown. This image is used to retrieve the contour and the silhouette of the ship. This silhouette is compared to a database containing different ships under different aspect angles, and the best five fits are presented to the user. The database is constructed using silhouettes from different open and closed sources.

3.5 Complex event detection

A final example is the detection of complex events. Here we devised a pipeline where, as a first step, relevant objects such as humans, vehicles and ships are detected and tracked. Features of these objects such as their behavior are also detected. In a second step, symbols are described which are logical operators that describe actions. In the last step complex level events are detected using user defined rules. The second and third part of the pipeline are reasoning frameworks. For this application we also designed interfaces that can be used to set the parameters for this reasoning.

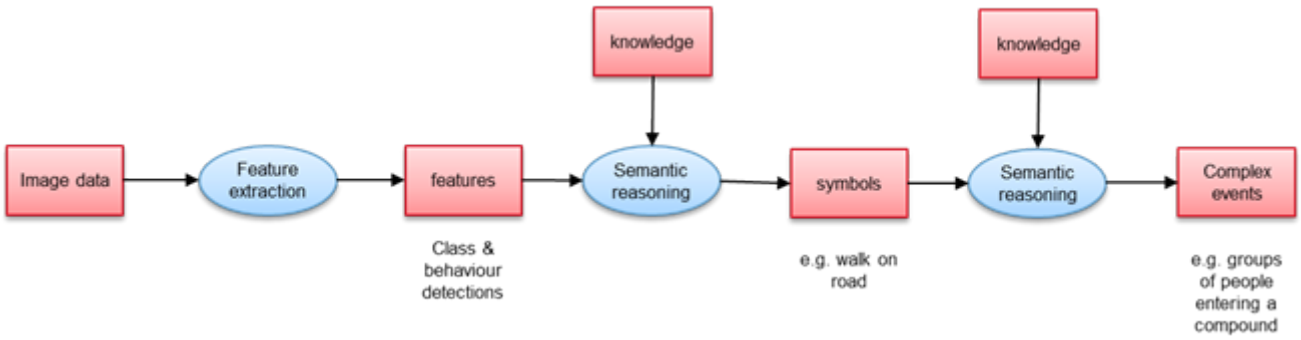

Figure 14 Design pattern for complex event detection

In [13][14], experiments for complex event detection were described, where this framework was applied for specified use cases. For these experiments video data was recorded including complex events such digging, climbing over a fence, crowd behavior, road blocks and drone attacks. At the first stage, features are computed such as class features (e.g., 'Weapon', 'Vehicle', 'Person', 'Car') or behavior features (e.g., 'Climb', 'Dig'). The output of the features are class or behavior detections, with a certain confidence.

Subsequently, symbols are created. Symbols are logical operators that describe short-term behavior (e.g., action recognition or track-based analysis), such as 'Walk on road' and 'Climb over the fence. These symbols are defined by a user. At the top level, sentences are computed that are high-level descriptions of behavior. Sentences combine multiple symbols in a specified temporal order to describe long-term behavior. These sentences are also defined by a user. The result is a system in which a user can online be alerted for high level events as well as search offline for the high-level events of interest. Examples of the interface for symbol and sentence selection are shown in Figure 15. An example of a detected complex event: a group of people entering a gate, is shown in Figure 16.

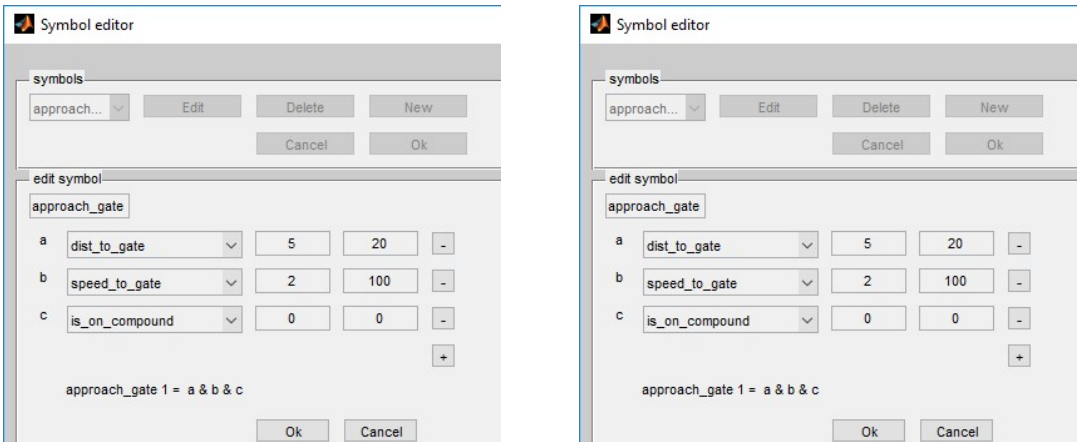

Figure 15 Example interface for symbol definition (left) and sentence definition (right).

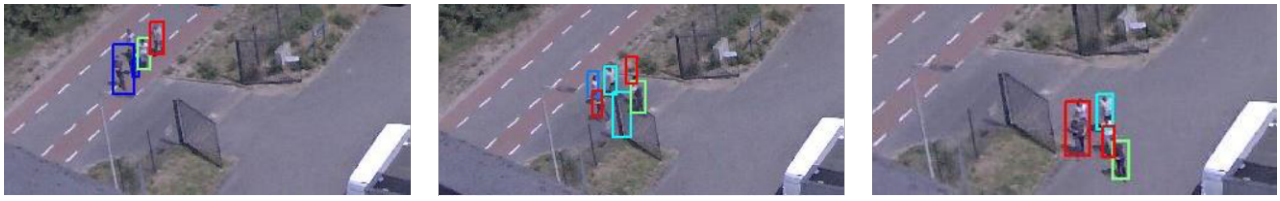

Figure 16 Example of complex event detection: a group of people entering a compound. At the left image, an approach of the gate is detected. In the center image, the entering of the gate can be seen and in the right image the persons are on the compound.

4. CONCLUSION, DISCUSSION AND FUTURE WORK

In this paper we have shown several Hybrid AI design patterns for different military imaging applications. It can be seen that a combination of data-driven learning and semantic reasoning is a sensible design option for such applications. In the future, we will further develop this approach with a more intertwined interaction between the reasoning and the data-driven part of the pipelines, and use more world knowledge, domain knowledge and relations between objects in the reasoning part. We will also look further into online learning, where the knowledge of the system performance will be used to ask the users relevant information.

Three of the examples focused on handling the challenge of too little training data. In these examples this was solved by data simulation, augmentation and transfer learning. This challenge will be further researched by focusing more on these approaches and develop possible new approaches such as online learning.

In the long term, our goals are to develop Hybrid AI algorithms that can 1) handle long term events (up to hours instead of seconds, 2) are context aware and adaptive to their environment, 3) can explain their reasoning or the quality of their results to a user and 4) can handle adversarial data, or in other words malicious inputs.

The added value of the architectural patterns for these approaches is that it provides a way to discuss the different elements in the processing pipeline. It stimulates discussion what the input and output of the different processing blocks are, and how they work together. Note that we used more detail in the different blocks to specify the type of data or algorithms that are applied. In future work we will investigate how elements such as online learning can be presented in this framework.

5. REFERENCES

[1] Van Harmelen, F., ten Teije, A., Compositional patterns for combining KR & ML: a first attempt", Pre-Proceedings of the Cognitive Computation Symposium: Thinking Beyond Deep Learning (CoCoSym 2018)

[2] Van Harmelen, Frank, and Annette ten Teije. "A Boxology of Design Patterns for Hybrid Learning and Reasoning Systems." *arXiv preprint arXiv:1905.12389* (2019).

[3] Serafini, L., & Garcez, A. D. A. (2016). Logic tensor networks: Deep learning and logical reasoning from data and knowledge. arXiv preprint arXiv:1606.04422.

[4] Maedche, A., & Staab, S. (2004). Ontology learning. In Handbook on ontologies (pp. 173-190). Springer, Berlin, Heidelberg.

[5] G. Burghouts, V1508: Full-Motion Video for Intelligence, Surveillance and Reconnaissance – Description of Algorithms for Metadata Extraction, TNO report R11164, 2018

[6] Anna News YouTube channel: https://www.youtube.com/channel/UCGib-bLlq8HTRp2YaEESxeg

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.- Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In European Conference on Computer Vision, pages 21–37. Springer, 2016. https://doi.org/10.1007/978-3-319-46448-0_2 (2000).

[8] Stap, N., van Opbroek, A. G., Huizinga, W., Wilmer, M. M. G., van den Broek, S. P., Pruim, R. H. R., ... & Dijk, J. (2018). Maritime detection framework 2.0: A new approach of maritime target detection in electro-optical sensors. Electro-Optical and Infrared Systems: Technology and Applications XV 2018, 12 September 2018 through 13 September 2018, Hickman, DL Bursing, H. Huckridge, DA, Proceedings of SPIE-The International Society for Optical Engineering, 10795.

[9] MIT. "Port of Single Shot MultiBox Detector to Keras". (2017) [Online]. Available: https://github.com/rykov8/ssd_keras. Accessed: August 2018.

[10] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., Zisserman, A. The PASCAL Visual Object Classes Challenge (VOC2007). http://www. pascal-network.org/challenges/VOC/voc2007/index.html.

[11] Van Ramshorst, A. (2018). Automatic Segmentation of Ships in Digital Images: A Deep Learning Approach.

[12] N. van der Stap, J. Dijk, V1423 MEOSS demonstration, TNO report R10308, 2018.

[13] Bouma, H., Schutte, K., ten Hove, J. M., Burghouts, G., & Baan, J. (2018, October). Flexible human-definable automatic behavior analysis for suspicious activity detection in surveillance cameras to protect critical infrastructures. In Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II (Vol. 10802, p. 108020N). International Society for Optics and Photonics.

[14] Schutte, K., Burghouts, G., van der Stap, N., Westerwoudt, V., Bouma, H., Kruithof, M., ... & ten Hove, J. M. (2016, October). Long-term behavior understanding based on the expert-based combination of short-term observations in high-resolution CCTV. In Optics and Photonics for Counterterrorism, Crime Fighting, and Defence XII (Vol. 9995, p. 99950Q). International Society for Optics and Photonics.