Issues and Approaches in Data Fusion

Peter Willett
University of Connecticut

peter.willett@uconn.edu

I. MARITIME ANOMALY DETECTION

A. Detecting Stealth Deviations from Standard Routes Using the Ornstein-Uhlenbeck Process

The Automatic Identification System (AIS) makes covert rendezvous at sea, such as smuggling and piracy, impossible. But, in practice, AIS can be spoofed or simply disabled. Here we report on [4], in which a novel anomaly detection procedure based on the Ornstein-Uhlenbeck (OU) mean-reverting stochastic process is presented. The considered anomaly is a vessel that deviates from a planned route as in figure 1, changing its nominal velocity v0. In order to hide this behavior, the vessel switches off its Automatic Identification System (AIS) device for a time T, and then tries to revert to the previous nominal velocity v0. The decision that has to be taken is either declaring that a deviation happened or not, relying only upon two consecutive AIS contacts. Furthermore, the extension to the scenario in which multiple contacts (e.g. radar) are available during the time period T is also considered. A proper statistical hypothesis testing procedure that builds on the changes in the OU process long-term velocity parameter v0 of the vessel is the core of the proposed approach and enables for the solution of the anomaly detection problem. Closed analytical forms are provided for the detection and false alarm probabilities of the hypothesis test.

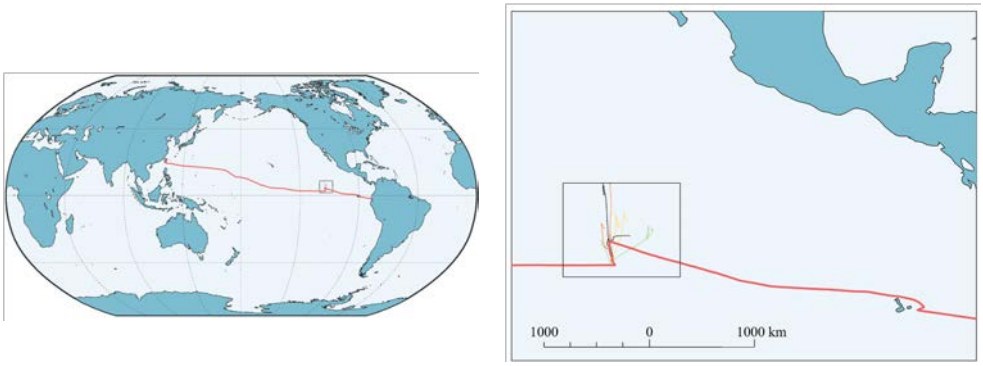

Fig. 1. The track of cargo vessel reveals a (possibly illegal) rendezvous with four fishing vessels in the Pacific Ocean. Left: world view. Right: close-up of diversion.

The OU model has a much tighter variance on predicted vessels' locations when they are out of view, hence is more suitable than (say) a nearly-constant velocity model that is more common for many tracking applications. When there are gaps in the data during which anomalies occur, these manifest themselves as significant excursions from the nominal velocity (and position). It will be seen that the GLRT for the hypothesis testing problem at hand can be easily traced back to the GLRT for Gaussian linear model [11], where the test statistics under the two hypotheses H_0 and H_1 are characterized, respectively, by a central and a non-central Chi-squared distributions, both with the same degrees of freedom. The performance is illustrated

in figure 2.

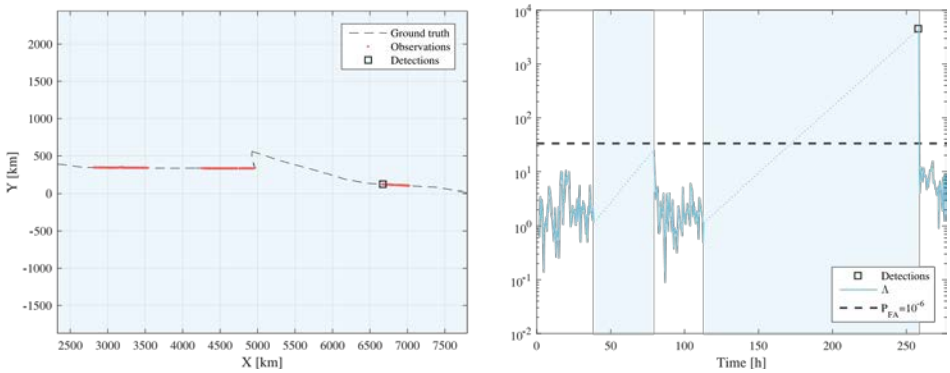

Fig. 2. For the example in figure 1 a gap in AIS coverage has been simulated (left) and the resulting test statistic (right) shows clearly the anomaly.

B. Malicious AIS Spoofing and Abnormal Stealth Deviations: A Comprehensive Statistical Framework for Maritime Anomaly Detection

The natural next question is whether a vessel can hide itself by issuing false AIS reports. The paper [5] assumes the vessel trajectory nominally follows a piecewise mean-reverting process; thereby, it addresses the problem of establishing whether a vessel is reporting adulterated position information through AIS messages in order to hide its current planned route and a possible deviation from the nominal route. Multiple hypothesis testing suggests a framework to enlist reliable information from monitoring systems (coastal radars and space-born satellite sensors) in support of detection of anomalies, spoofing, and stealth deviations. The proposed solution involves the derivation of anomaly detection rules based on the generalized likelihood ratio test and the model-order selection methodologies. The effectiveness of the proposed anomaly detection strategy is tested for different case studies within an operational scenario with simulated data.

Optimal Opponent Stealth Trajectory Planning based on an Efficient Optimization Technique

The previous subsection asks whether a vessel engaged in a course that it would prefer kept hidden can mask what it is doing by obfuscating the observations. The next step in the “game” of spotting/evading maritime anomalies, also playing the opponent’s side [1], is to describe the least-detectable trajectory that that the elusive vessel can take. The opponent’s route planning problem is formalized as a non-convex optimization problem capitalizing the Kullback-Leibler (KL) divergence between the statistical hypotheses of the nominal and the anomalous trajectories as key performance measure. The velocity of the vessel is modeled with an Ornstein-Uhlenbeck (OU) mean reverting stochastic process, and physical and practical requirements are accounted for by enforcing several constraints at the optimization design stage. To handle the resulting non-convex optimization problem, we propose a globally-optimal and computationallyefficient technique, called the Non-Convex Optimized Stealth Trajectory (N-COST) algorithm. The NCOST algorithm consists of solving multiple convex problems, where the number was proportional to the number of segments of the piece-wise OU trajectory. The effectiveness of the approach proposed is demonstrated through case studies and a real-world example, see figure 3. The game can be extended to the detector specifically looking for the least-observable trajectory.

The main contribution of [1] is twofold, specifically: i) a novel optimal path planning formulation to blind an anomaly detection procedure so as to execute covertly a given task (e.g., a rendezvous with

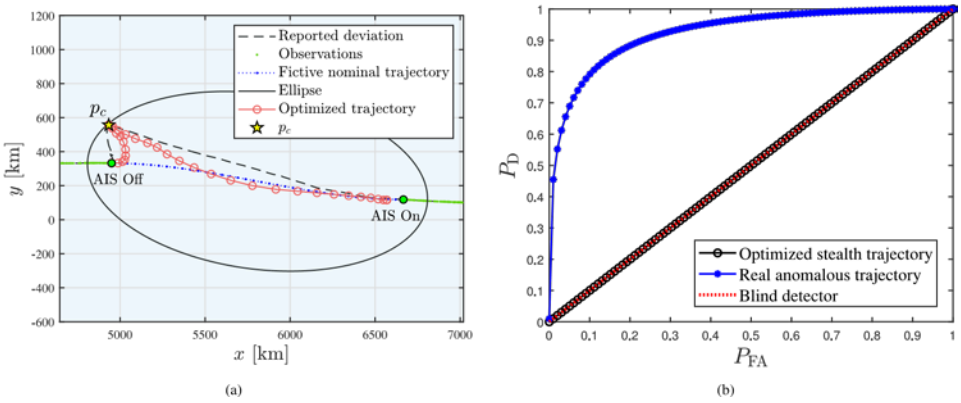

Fig. 3. Real-world AIS track (black dashed line) versus optimized trajectory (red circle markers line). (b): Performance of anomaly detector (10) shows the optimized trajectory being less detectable than the real one.

another ship), and ii) an optimal and efficient solution for the aforementioned problem formulation, called Non-Convex Optimized Stealth Trajectory (N-COST). Without the (non-trivial) derivation of N-COST, the proposed optimization problem would not have been solvable. A brute force implementation procedure would not be feasible, since the optimization space would grow exponentially with the length of the velocity and position sequence for any given discretization step of the optimization variables.

II. MARITIME TRAJECTORY LEARNING

A. Multiple Ornstein-Uhlenbeck Processes for Maritime Traffic Graph Representation

In [8] we propose an unsupervised procedure to automatically extract a graph-based model of commercial maritime traffic routes from historical Automatic Identification System (AIS) data. In the proposed representation, the main elements of maritime traffic patterns, such as maneuvering regions and sea-lanes, are represented, respectively, with graph vertices and edges. Vessel motion dynamics are defined by multiple Ornstein-Uhlenbeck processes with different long-run mean parameters, which in our approach can be estimated with a change detection procedure based on Page's test, aimed to reveal the spatial points representative of velocity changes. A density-based clustering algorithm is then applied to aggregate the detected changes into groups of similar elements, and to reject outliers. To validate the proposed graphbased representation of the maritime traffic, two performance criteria are tested against a real-world trajectory dataset collected off the Iberian Coast and the English Channel. Results show the effectiveness of the proposed approach, which is suitable to be integrated at any level of a JDL system.

A well-known fact about maritime traffic is that the majority of it is, unsurprisingly, very regular, as also shown on the left in figure 4, where a ship traffic density map is reported. Ships, especially those involved in goods transportation, often if not always seek to optimize fuel consumption, and therefore will naturally follow the most convenient path allowed by international regulations and traffic separation schemata. Thanks to this overall regular behavior, detections of abrupt changes in the long-run velocity parameter v are more likely expected in proximity of waypoint areas. Formalized via change detection theory [12] the procedure used aims to detect these changes and subsequently to estimate the time instants of change via a Page test. In order to discover significant waypoint areas, i.e., regions where vessels are more likely to show long-run mean velocity changes, the change detections need to be first aggregated in clusters. In this paper, we use the Density-Based Spatial Clustering of Applications With

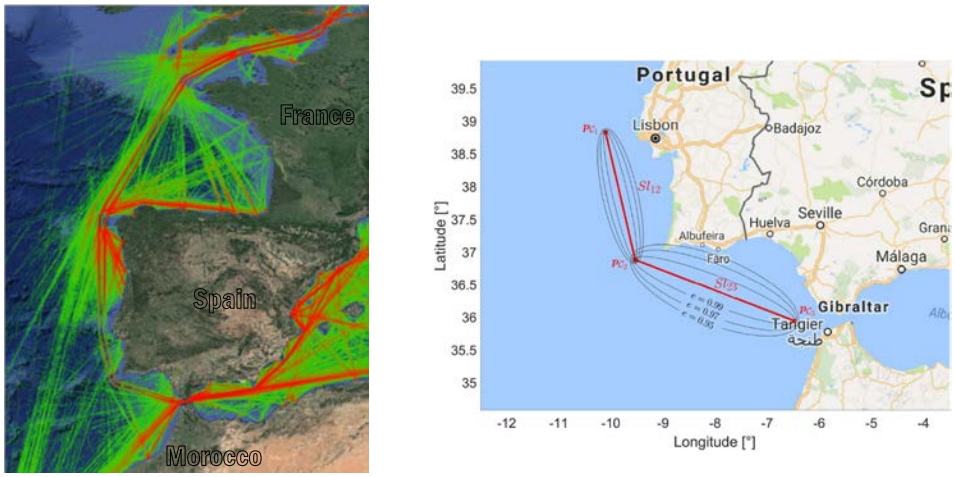

Fig. 4. Left: Density map of AIS messages collected from cargo ships during two months in the area of interest. Several waypoints, i.e., localized spatial regions where vessels show velocity or orientation changes are observable. Linear piecewise lanes that make up standardized routes followed by vessels are clearly visible. The color is proportional to the number of ships whose reported positions fall within the ground footprint, from green (low traffic density) to red (high traffic density). Right: Two navigational legs off the coast of the Iberian peninsula and elliptic areas defined by three different values of eccentricity: 0.95, 0.97, 0.99.

Noise (DBSCAN) algorithm [10]. An example is shown on the right of 4. The ellipses are used in the clustering procedure.

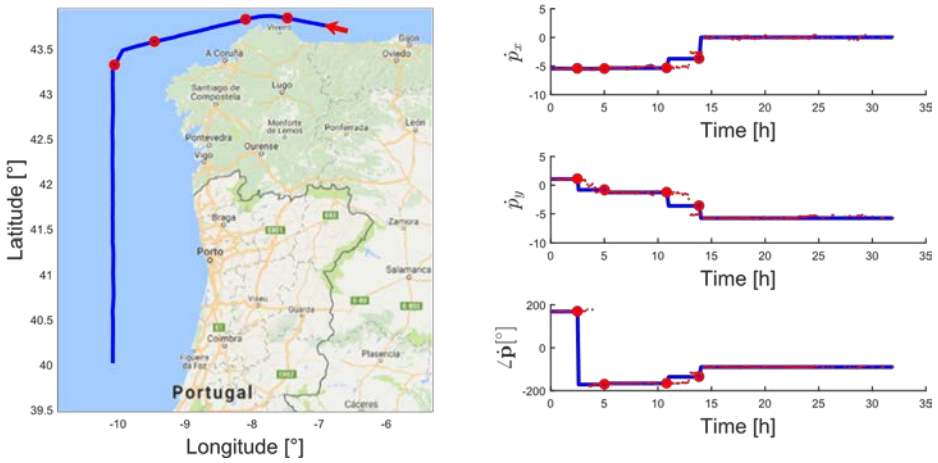

Fig. 5. Example of the change detection procedure applied to a vessel's track. On the left, the red circles are the points where a change is detected, while the red arrow represents the starting point and the vessel's direction. On the right, the red dots represent the instantaneous values of the velocity along the x and y components and the track orientation while the circles detected changes of the parameters. The parameter δ of the change detection procedure is set to 5^o and N = 15.

B. Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks

Recent deep learning methods for vessel trajectory prediction are able to learn complex maritime patterns

from historical Automatic Identification System (AIS) data and accurately predict sequences of future vessel positions with a prediction horizon of several hours. However, in maritime surveillance applications, reliably quantifying the prediction uncertainty can be as important as obtaining high accuracy. The paper [7] extends deep learning frameworks for trajectory prediction tasks by exploring how recurrent encoder-decoder neural networks can be tasked not only to predict but also to yield a corresponding prediction uncertainty via Bayesian modeling of epistemic and aleatoric uncertainties. We compare the prediction performance of two different models based on labeled or unlabeled input data to highlight how uncertainty quantification and accuracy can be improved by using, if available, additional information on the intention of the ship (e.g., its planned destination).

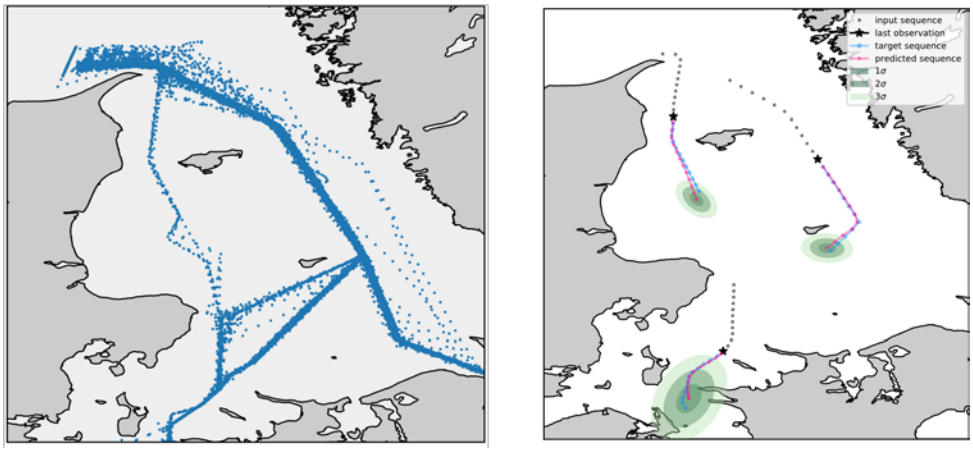

Fig. 6. Left: Complete dataset of AIS positions used in the experiments; the training and testing sets are subsets of the dataset showed here. Each dot in the image corresponds to a ship's position. Contains data from the Danish Maritime Authority that is used in [7] in accordance with the conditions for the use of Danish public data [9]. Right: Vessel trajectory prediction with representation of the prediction uncertainty depicted as confidence ellipses (at different confidence levels), using [9].

The main contributions of [7] are: 1) A model for the aleatoric and epistemic uncertainty of trajectory predictions provided by encoder-decoder RNNs using Bayesian deep learning tools; 2) A novel regularization method for the decoding phase to prevent complex co-adaptations between the encoded data and the high-level information about the vessel's intention; and 3) Experimental results on real-world AIS data showing the effectiveness of the proposed encoder-decoder architecture with uncertainty modeling in learning trajectory predictions with well-quantified uncertainty estimates using labeled and unlabeled data. Figure 6 shows an example of how the proposed model is able to predict future trajectories and prediction uncertainty given input sequences from past data.

Uncertainty on the prediction estimates can be captured with recently developed Bayesian deep learning tools, which offer a practical framework for representing uncertainty in deep learning models. In the context of supervised learning, two forms of uncertainty, i.e., aleatoric and epistemic uncertainty are considered, where epistemic is the reducible and aleatoric the irreducible part of uncertainty. Aleatoric (or data) uncertainty captures noise inherent to the observations, whereas epistemic (or model) uncertainty accounts for uncertainty in the neural network model parameters. Epistemic uncertainty is a particular concern for neural networks given their many free parameters, and can be large for data that is significantly different from the training set. Thus, for any real-world application of neural network uncertainty

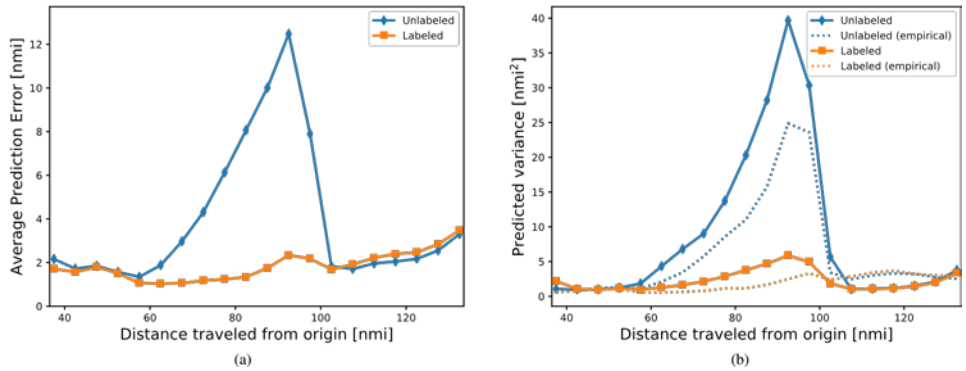

Fig. 7. Average Prediction Error (APE) (a) and predicted total variance (b). Both are computed at the h^{th} prediction sample (i.e., prediction horizon 3 hours) and are expressed as a function of the vessel distance from a fixed point, located in the upper left corner of figure 6. Panel (b) shows the square root of the determinant of the covariance matrix produced by the network, which is also known as generalized variance and is proportional to the area of the prediction uncertainty ellipse.

estimation, it is critical that it be taken into account. We follow a combined aleatoric-epistemic model to capture both aleatoric and epistemic uncertainty in our prediction model. We use an encoder-decoder architecture with attention mechanism composed of a BiLSTM encoder layer with 64 hidden units, and an LSTM decoder layer with 64 hidden units. For Bayesian modeling of the epistemic uncertainty, we used MC dropout with M = 100 samples, and dropout rate applied to recurrent connection p = 0.05 in both encoder and decoder layers. The model was trained by applying AdamW optimizer with a learning rate of 0.0001 and weight decay of 0.0001 to minimize the mean absolute error loss function, an earlystopping rule with 3000-epoch patience, and a mini-batch size of 200 samples. Better performance in terms of prediction accuracy can be achieved by labeling the input data based on a high-level pattern information, such as the vessel's intended destination. Figure 7 demonstrates both the effectiveness of our scheme to model uncertainty and also the impact of labeling (e.g., of destination) on training data.

III. SPACE-BASED GLOBAL MARITIME SURVEILLANCE

Maritime surveillance (MS) is of paramount importance for search and rescue operations, fishery monitoring, pollution control, law enforcement, migration monitoring, and national security policies. Since terrestrial radars and automatic identification system (AIS) are not always guaranteed to provide a comprehensive and seamless coverage of the entire maritime domain, the exploitation of space-based sensor technologies installed on satellites orbiting around the Earth, that complements existing terrestrial techniques, is crucial. These space-based technologies (see figure 8) include satellite AIS, synthetic aperture radars, multi-spectral and hyperspectral optical sensors, and global navigation satellite system reflectometry, and are reviewed in [2]. Furthermore, the development of future MS systems combining multiple terrestrial and space-based sensors with other information sources requires dedicated artificial intelligence and data fusion techniques for the processing of raw satellite images and fuse heterogeneous information. The objective of [3] is to provide an overview on the most promising artificial intelligence and data fusion techniques for MS using space-based sensors.

A. Satellite Technologies

Nowadays, space-based remote sensing data involves a range of technologies and modalities including satellite AIS (Sat-AIS) (see figure 9), synthetic aperture radar (SAR) (see figure 10), multi-spectral (MSP)

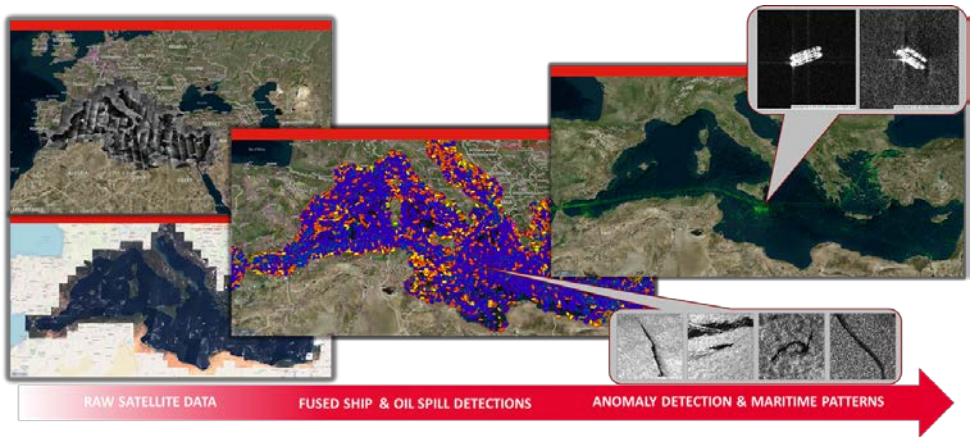

Fig. 8. The processing of images acquired by space-based sensors with state-of-the-art artificial intelligence and data fusion techniques to extract real-time information on the current maritime situation enables to detect and prevent piracy, human trafficking, and environmental threats such as oil spills.

and hyper-spectral (HSP) optical sensors, Global Navigation Satellite System reflectometry (GNSS-R), and characteristics, e.g., resolutions, viewing angle, frequency, acquisition modes, polarization. Space-based sensors for Earth observation (EO) installed on artificial satellites allow collection of images of huge and remote areas of the globe within relatively short times, and hence highly important for MS improvement [2]. Recently, different initiatives from national and international space agencies have encouraged the exploitation of remote sensing technologies by providing end-users with freely available data. The joint NASA/USGS Landsat series and the joint EU/ESA Copernicus program with the Sentinel multisensor constellation for EO and monitoring, the NASA CYGNSS mission for the hurricane forecasts and ocean surface analysis using GNSS signals of opportunity, represent a partial list of space programs and missions providing satellite data at no cost.

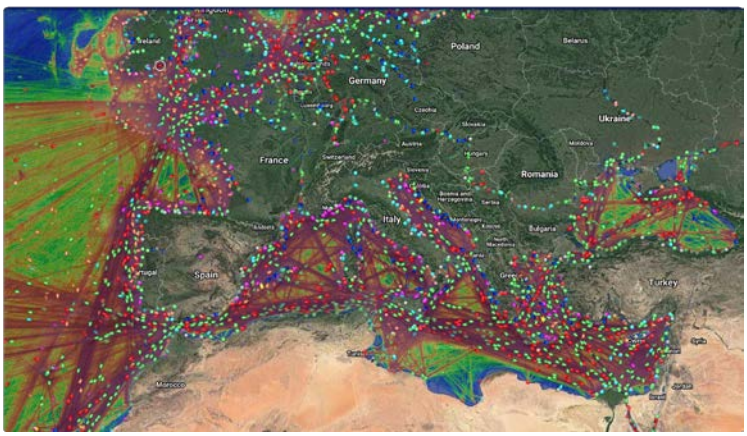

Fig. 9. Real-time AIS reports in the Mediterranean Sea, acquired by ground-based stations and satellites (image courtesy of MarineTraffic).

Each of the technologies mentioned has advantages, but also some corresponding concerns. For example, GNSS-R offers a highly-promising way to harness GPS signals that are extant; but it is an unproven

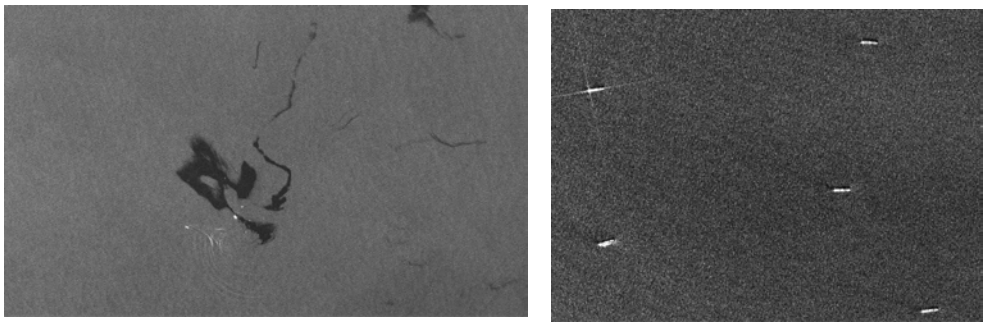

Fig. 10. Left: Oil spill detection on a SAR image acquired by S-1 constellation in IW mode (10m resolution). Right: Vessels detected using a SAR image acquired by COSMO-SkyMed constellation in Stripmap HIMAGE mode (5m resolution).

Technology	Advantages	Drawbacks
Sat-AIS	Accurate ship information Very high update rate (up to 2 seconds) Global coverage of AIS services Compact, low-power, light-weight and cheap	Vulnerable (e.g. spoofing) Limited to ships with AIS Limited to cooperative ships Limited to ship detection/tracking
SAR	All-weather, all-time sensing capabilities Very high spatial resolution (down to 1 m) Polarimetric diversity	Large size Imagery difficult to visually and manually interpret Affected by speckle noise
MSP	Imagery easy to interpret Very high spatial resolution (down to 0.3 m)	Sensitive to cloud and sunlight conditions Limited revisit time Limited areas covered during each acquisition
HSP	Very high spectral resolution (down to 1 nm) Anti-camouflage capabilities Suited to accurate classification	Low spatial resolution Large computational burden Sensitive to cloud and sunlight conditions
GNSS-R	All-weather, all-time sensing capabilities Compact, low-power, light-weight and cheap Seamless global coverage 100+ GNSS satellites available	Low spatial resolution Very low power density Poor performance in standard configuration Predetermined waveform

Fig. 11. Capabilities and disadvantages of the various satellite maritime surveillance technologies.

technology, and the power budget is of concern. A table listing the strengths and weaknesses of each is in figure 11.

B. Artificial Intelligence and Dats Fusion

In [3] we describe the main artificial intelligence (AI) and imaging techniques for image segmentation, target detection and classification, and provide possible use cases with real images acquired by satellite sensors. We then proceed by describing the most recent Bayesian and statistical techniques to extract valuable knowledge from a huge amount of historical Sat-AIS data, such as most common maritime routes, and to track multiple targets by fusing information collected by multiple heterogeneous sensors. Among these, multitarget tracking (MTT) algorithms based on the sum-product algorithm (SPA) are gaining popularity thanks to their abilities to fuse information from different heterogeneous sources, to their scalability, i.e., low computational complexity at the increasing number of information sources, targets and measure-=ments, and

to their capability at including contextual information, e.g., maritime routes and ships class information extracted from satellite images. A use case that confirms the strength of SPA-based MTT algorithms when combined with information acquired by satellite sensors is also provided.

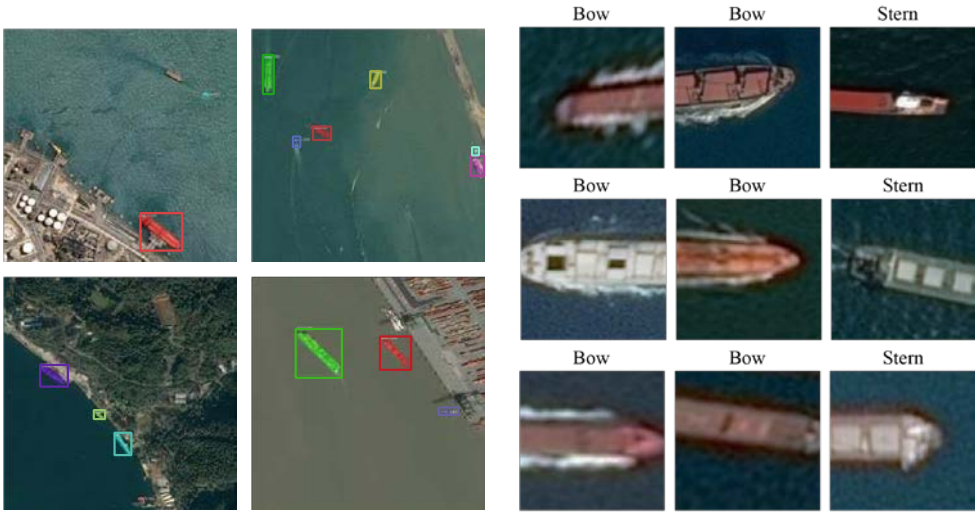

Fig. 12. Left: Example result of the segmentation task on VHR images (first use case): detected ships are surrounded by bounding boxes and highlighted with different colors. Right: Stern and bow classification for detected ships in clipped satellite images acquired by S-2.

The training of complex DNN architectures with multiple hidden layers and neurons for satellite image segmentation and target classification requires the avail- ability of large labelled datasets, which need to be pre- processed, in order to reduce unwanted bias in the data, and the availability of high computational power, such as graphical processing units (GPUs). Data augmentation techniques, e.g., cropping, padding, and flipping, enable practitioners to significantly increase the diversity of data images available to train large DNN models, without actually collecting new images. One use case consists of detecting ships and extracting their important features, e.g, position, width, length, heading and other relevant information, from very high resolution (VHR) optical satellite images by means of CNNs. The training dataset consists of about 200000 VHR images, with spatial resolution of approximately 1.5 m and of dimension 768 by 768 pixels, acquired by GeoEye, SPOT, Pleiades and Black Sky satellites; the inference, i.e., the testing of the CNN, is performed on new images acquired by these sensors. A Mask R-CNN architecture is used, and the left panel in figure 12 shows an example result of the segmentation task: detected ships are surrounded by bounding boxes and highlighted with different colors. Another use case is related instead to the classification between stern and bow of ships by means of a ResNet34 architecture. The training dataset is composed of satellite images acquired by Sentinel-2 (S-2); detected ships are clipped, manually divided into stern and bow, and labelled accordingly. The right panel in figure 12 represents the results of the classification.

The increasing availability of space-based remote sensors providing continuous coverage on remote areas of the Earth requires the development of sophisticated information fusion and MTT algorithms, which are able to combine and fuse information from different heterogeneous sources, e.g., terrestrial radars, Sat-AIS, SAR and optical sensors. One use case is therefore to fuse SAR measurements and AIS messages in order to track potential targets and also identify them by exploiting the MMSI identifiers. The SAR measurements and AIS messages used by the SPA-based MTT method are represented in the left panel of figure 13 by the red crosses and green triangles, respectively. The red dashed rectangle delimits a geographical area of little interest from a tracking point of view with a large number of docked

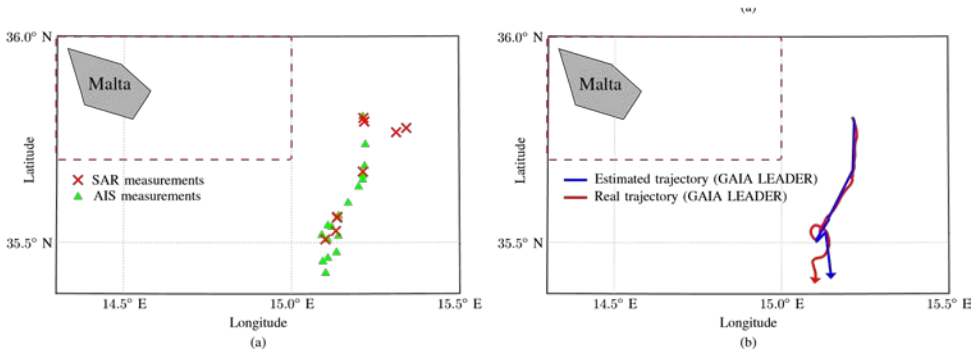

Fig. 13. Left: SAR (red crosses) and AIS (green triangles) measurements used by the SPA-based fusion algorithm obtained in the period from 8 September 2018 at 16:30 till 12 September 2018 at 16:47. The measurements inside the red dashed rectangle have not been considered. Right: Estimated and real trajectories for the vehicles carrier Gaia Leader (MMSI 371158000) obtained using the SAR and AIS measurements.

ships. Therefore, the SAR measurements and AIS messages from ships within this red dashed rectangle are discarded. The dynamic model used to describe ship motion in the SPA-based MTT algorithm is the OU model. Estimated positions of detected targets are calculated each time a SAR measurement is obtained, using all AIS messages gathered in the time interval between the current and the previous SAR measurements. Besides estimating the number of targets and their positions, the algorithm associates, when possible, an MMSI identifier to each of them. These MMSI identifiers are selected from a set consisting of all the MMSI identifiers so far observed. The right panel of figure 13 shows the output of the SPA-based MTT algorithm. One can see that in the considered time period only a single target is detected. In particular, the red and blue lines show the true and estimated trajectories of the vehicle carriers GAIA LEADER (see figure 14).

Fig. 14. The vehicles carrier GAIA LEADER (MMSI: 371158000). (Image courtesy of MarineTraffic.com).

REFERENCES

[1] A. Aubry, P. Braca, E. d'Afflisio, A. De Maio, L. Millefiori and P. Willett, "Optimal Opponent Stealth Trajectory Planning Based on an Efficient Optimization Technique," *IEEE Transactions on Signal Processing*, pp. 270-283, December 2020.

[2] G. Soldi, D. Gaglione, N. Forti, A. Di Simone, F.-C. Daffina, G. Bottini, D. Quattrociocchi, L. Millefiori, P. Braca, P.` Willett, A. Iodice, D. Riccio and A. Farina, "Space-based Global Maritime Surveillance – Part 1: Satellite Technologies," *IEEE AES Systems Magazine*, pp. 8-28, September 2021.

[3] G. Soldi, D. Gaglione, N. Forti, A. Di Simone, F.-C. Daffina, G. Bottini, D. Quattrociocchi, L. Millefiori, P. Braca, P.` Willett, A. Iodice, D. Riccio and A. Farina, "Space-based Global Maritime Surveillance – Part 2: Artificial Intelligence and Data Fusion Techniques," *IEEE AES Systems Magazine*, pp. 30-42, September 2021.

[4] E. d'Afflisio, P. Braca, L. Millefiori and P. Willett, "Detecting Anomalous Deviations from Standard Maritime Routes Using the Ornstein-Uhlenbeck Process," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6474-6487, December 2018.

[5] E. d'Afflisio, P. Braca and P. Willett, "Malicious AIS Spoofing and Abnormal Stealth Deviations: A Comprehensive Statistical Framework for Maritime Anomaly Detection," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 2093-2108, August 2021.

[6] S. Capobianco, N. Forti, L. Millefiori, P. Braca and P. Willett, "Uncertainty-Aware Recurrent Encoder-Decoder Networks for Vessel Trajectory Prediction," *Proceedings of the ISIF Fusion Conference*, South Africa, November 2021.

[7] S. Capobianco, L. Millefiori, N. Forti, P. Braca and P. Willett, "Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks," *IIEEE Transactions on Aerospace and Electronic Systems*, pp. 4329-4346, December 2021.

[8] P. Coscia, P. Braca, L. Millefiori, F. Palmieri and P. Willett, "Maritime Traffic Representation using Multiple OrnsteinUhlenbeck Processes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 5., pp. 2158-2170, October 2018.

[9] "Data from the Danish AIS system," https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/default.aspx, accessed: 2020-07-09.

[10] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226-231.

[11] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Volume 2, Prentice-Hall, 1998.

[12] E. Page, "Continuous inspection schemes", *Biometrika*, vol. 41, no. 1/2, pp. 100-115, June 1954.