ArtIndet

EUROPEAN DEFENCE AGENCY	*Service Framework Contract for the provision of a Study for "Artificial Intelligence for Automatic Detection Recognition, Identification (DRI) and Tracking - (ARTINDET)"* **Executive summary**	Aerospace and Defense everis

Summary: Executive summary

Record of Modifications

Version	Description [or description of changes]	Author	Date of Creation	Approved by	Date of Approval
1.0	First version	everisADS	09/11/2021		
2.0	Second version	everisADS	02/12/2021		

INDEX

1. Introduction

This project analyses new techniques of imagery-based systems relying on Artificial Intelligence (AI) by including some techniques based on the deep learning paradigm. For that aim, images captured by high-resolution cameras and operating with Wide Field of View (WFOV) feed the different algorithms tested. In addition, applications for Unmanned Air Vehicle (UAV) mounting high-resolutions imagery cameras cannot stream all the gathered information to the ground station in real-time. Therefore, embedded systems are crucial to allow the feasibility of the implemented technology. The outputs of the study consist of:

- Selecting, describing and analysing the most promising AI-based automatic DRIT techniques.
- Clearly and quantitatively showing the improvement in term of performances in specific selected military use cases also emphasizing the added value for the final users for the situational awareness interpretation.
- Define a roadmap for results exploitation aiming at defining new project initiatives and at developing new EU defence capabilities.

2. Objectives

The general objective of this study is to analyse the most promising AI-based techniques for Automatic Detection Recognition Identification and Tracking (A-DRIT) at least in the following operative conditions:

- Spatially distributed small targets in strong background.
- High resolution imaging systems (passive or active) with wide Field of View (Wide region surveillance).
- Multidimensional imagery (multiband/ multispectral/hyperspectral, multiresolution).
- Multiple cooperating platforms.

The specific objectives of the study are:

- Identify which AI techniques can be used for implementing A-DRIT and provide the new related image processing system schemes.
- Select some specific military use cases where AI can provide tangible benefits.
- Make a performance analysis of the selected techniques in the specific identified use cases relevant to the above listed operative conditions, in comparison with the current most common DRIT techniques.
- Provide a clear roadmap about exploitation of results for new Cat B project and new EU defence capabilities.

3. Methodology

The different steps implemented in the project define a methodology based on AI for military purposes that has been summarized in Figure 1.

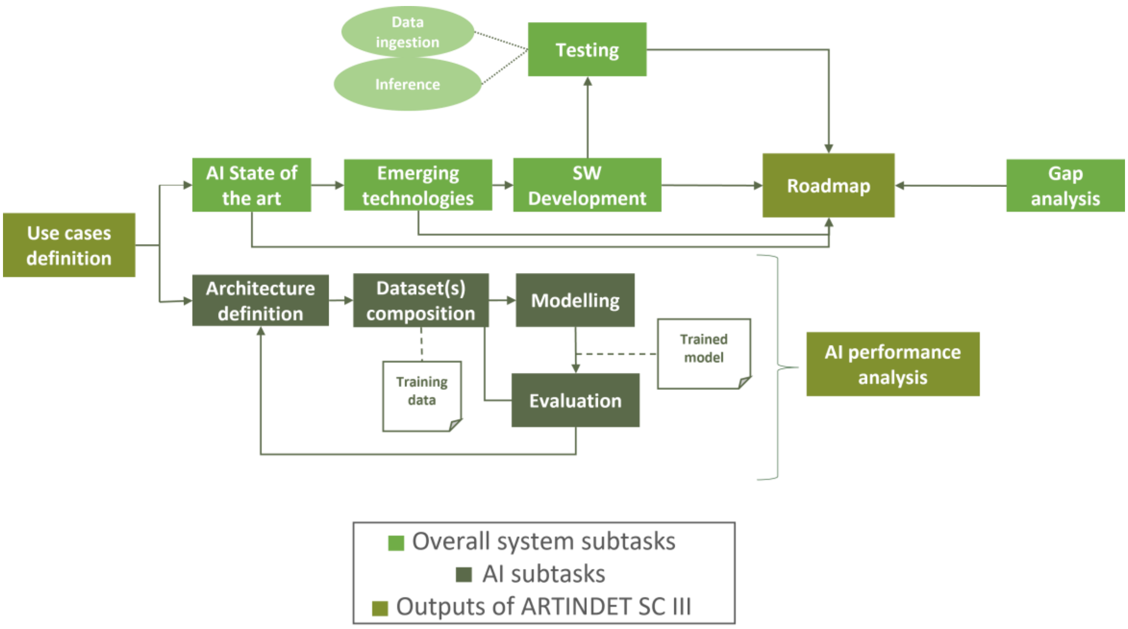

Figure 1 Methodology followed in ARTINDET

4. Project results

The project has gathered results in four areas:

- **Military use cases** (D1 RPAS operative missions and technical requirements)
- **AI methodology** (*D2 AI techniques selection and analysis for RPAS platforms*)
- **Comparison of selected AI techniques** (*D3 Performance analysis of the implemented AI based integrated technique*)
- **Roadmap** (*D4 Results exploitation and roadmap definition*)

4.1 Military Use cases

Among the use cases, two have been selected for development: monitoring and surveillance. The first selected use case is located in an urban or road scenario, and the second may be located in a maritime scenario. The platform selected for both use cases is the most widely used: Small UAV. Regarding the sensing devices, the SAR imaging sensing device complemented by an EO camera is present in both use cases. The overview of the chosen use cases can be found in Table 1.

	MISSION	SCENARIO	PLATFORM	SENSOR	TARGET
UC1	Monitoring	URBAN / ROAD	Small UAV	EO/SAR	Urban changes
UC2	Surveillance	MARITIME	Small UAV	EO/SAR	Ships

Table 1 Use cases specification

4.2 AI methodology

Different natures of datasets have been explored for all the stages included in the A-DRIT system. It shall be remarked the lack of available military datasets, especially in the SAR domain where the generic available datasets are less frequent. For this reason, this study is performed through satellite images. The satellite images are comparable though.

UC1 datasets are used for segmentation and UC2 datasets are also used for segmentation and detection. An overview of the selected datasets, classified by data type and use case, can be found in Table 2.

	UC1	UC2
EO	SpaceNet 6 Inria Aerial Label	AirBus
SAR	SpaceNet 6 Terra SAR-X	HRSID SSDD Terra SAR-X

Table 2 Summary of the selected datasets for this work.

The methodology proposed to carry out this task consisted of a "systematic review". This well-defined methodology was used to identify, analyse and interpret all available evidence related to a situation to be analysed in a comprehensive, unbiased, (to some extent) repeatable and highly scientifically valuable way. Systematic analysis is divided into three stages that are carried out iteratively: (1) Planning the analysis; (2) Conducting the analysis; and (3) Dissemination. In this task, the main stage was the conduct of the review, as the planning was done beforehand, and the dissemination is done at the end of the project.

The AI techniques used for the different use cases can be found in Table 3.

	UC1	UC2
Technique	Segmentation	Segmentation Detection

Table 3 Summary of the techniques applied to each use case

The selected techniques are segmentation and detection and both of them find automatically locations of certain items of interest in an image. However, these techniques are different and thus, the architecture of the convolutional neural network that perform these tasks is also different, which implies that the representation of the outcome is also different. They can take care of the whole A-DRIT chain as the differences in the representation of the outcomes do not impact in these objectives. A basic way of tracking is following their outcomes.

In the case of detection, it takes an image as input and produces one or more bounding boxes with the class label attached to each bounding box. These algorithms are usually capable to deal with multi-class classification. There are different architectures that perform detection through bounding boxes but in this project the adopted solution is Yolo v5.

Image segmentation is a further extension of object detection in which the presence of an object is marked through pixel-wise masks generated for each object in the image, i.e., the segmentation mask. This technique is more granular than bounding box generation because it determines the shape of each object present in the image. Unet with EfficientNet as encoder is the architecture applied for segmentation in this work.

4.3 Comparison of selected AI techniques

The selected AI techniques were implemented with the objective of comparing their performance, taking into account the operational missions of RPAS for A-DRIT purposes and the established technical requirements. To achieve this objective, the following steps were performed:

I. Data generation

Composition of the dataset to provide a comprehensive analysis and assessment of the different techniques. Limits related to the quality and size of the training and evaluation data were taken into account in order to define a suitable dataset. Datasets available online that met the above characteristics.

II. Definition of relevant Key Performance Indicators (KPIs)

- Definition of relevant KPIs to compare AI-based A-DRIT with current DRIT techniques in the selected use cases. These indicators will help during the comparison process and the pros and cons analysis.

- Training procedure and plan and comparative performance evaluation and analysis in the selected use cases.
- Evaluation of training parameters.
- Generation of evaluation tests.

III. Pros and cons analysis

Identification of the strengths and weaknesses of the compared A-DRIT technologies based on the previously evaluated KPIs.

4.3.1 UC1

For the use case 1, the recognition of infrastructures, the UNet neural network for semantic segmentation is employed, presented in Section 4.1. This architecture, which was develop for the SpaceNet 6 competition, required some adaptations to match the objectives set for this project.

The developments regarding this architecture in this project have been aimed to improve the efficiency of the neural network, by using a single instance of it, and to make it more generic, allowing the use of different datasets.

Respecting the datasets employed to obtain the results for this UC1, SpaceNet, INRIA and TerraSAR-X were used. The first, which includes several imaging modes and spatially synchronized crops, have been further examined as it allows to test different combinations of input data in the same grounds. The second, that contains more diverse locations, is used to confirm the EO results obtained in SpaceNet. And the third, is used to analyze the influence of having less polarization data, as the SAR images only contain one channel against the four different polarizations of SpaceNet.

4.3.2 UC2

For the use case 2, the detection of boats, two different neural networks are employed, the UNet and Yolo architectures, both presented in Sections 4.1 and 4.2. Due to being architectures that perform different tasks, semantic segmentation and detection respectively, most of the work went in to adapting the datasets for each of them so the same data can be loaded and compared. Additionally, the UNet network required some changes to move from buildings to segmenting boats.

Respecting the datasets employed to obtain the results for UC2, Airbus Ship Detection dataset, HRSID and TerraSAR-X were used. The first, is used to provide the EO data, whereas the second and third, contain SAR information. In this case, contrary to the datasets in UC1, the SAR datasets only contain the polarimetric information of one channel, so the influence of having more detailed data is out of the scope.

4.4 Roadmap

The roadmap built has been divided in two terms:

- Short-term
- Mid-term

The first is located on a lower TRL (3-5), and its objective is to perform a proof of concept. The second is located on a higher TRL (4-6) and aims to cover all the functionalities and perform field tests.

The step challenges defined for them are enumerated below:

- Short-term
  - Dataset definition and classification in the military domain
  - Applying supervised AI for target detection and tracking in military scenarios
  - Real unmanned systems with AIDRIT capabilities
- Mid-term
  - Applying supervised AI for target detection and tracking in military scenarios
  - Real unmanned systems with AIDRIT capabilities

5. Conclusions

The study performed in ARTINDET has provided a more precise knowledge regarding the feasibility and the steps required to adopt A-DRIT systems in the military domain. The project has gone through all the stages required for this kind of analysis:

- **State of the art**
- **Use case definition**
- **Artificial Intelligence emerging technologies and techniques**
- **Performance analysis**
- **Exploitation and roadmap**

The main outcomes of the project are:

- **Definition of two use cases**

  Two use cases (UC1 and UC2) are defined base on stakeholder consultation. UC1 is focused on the detection and recognition of critical buildings and infrastructures subject to suffer a terrorist attack based on the imaging data provided by a satellite, while UC2 is defined on the Ships segmentation, where the sources analysed select those areas at the port with tremendous cargo handling capacity or the crisscrossed busy canals throughout the trading cities. These areas can simultaneously present specific scenes in need with the limited swath in high-resolution SAR imageries. The summary of the use cases can be noticed in the next table.

	MISSION	SCENARIO	PLATFORM	SENSOR	TARGET
UC1	Monitoring	URBAN / ROAD	Small UAV	EO/SAR	Urban changes
UC2	Surveillance	MARITIME	Small UAV	EO/SAR	Ships

Table 4 4 Outputs of ARTINDET SC3: Use cases

- **Dataset**

  It is discussed in section 4.2, where it can be seen how the lack of drone imagery of the use cases can be a problem, but satellite imagery datasets were used, which are comparable to drone imagery. On the other hand, EO and SAR imagery must be synchronised in their geolocation. The summary of the constructed datasets can be seen in the next table.

	UC1	UC2
EO	SpaceNet 6 Inria Aerial Label	AirBus
SAR	SpaceNet 6 Terra SAR-X	HRSID SSDD Terra SAR-X

Table 5 Outputs of ARTINDET SC3: Datasets.

- **Selection of most relevant techniques of AI and performance analysis**

  This work presents a complete analysis of the performance of the systems defined in D2: AI techniques selection and analysis for RPAS of the project ARTINDET SC3 Artificial Intelligence for Automatic Detection Recognition, Identification (DRI) and Tracking. The uses cases studied were selected in the aforementioned document and they are building /Infrastructure localization and ship detection. Those use cases require a special acquisition procedure to cover the AOI with the sensing equipment defined (EO and SAR) and a specific software architecture to host the methods needed.

BEST HYPERPARAMETERS COMBINATION			
Use Case	**Architecture**	**Learning Rate**	**Batch Size**
1	UNET	0,0001	16
2	UNET	0,0001	8
	YOLO	0,0010	8

Table 6 *Outputs of ARTINDET SC3:* Best hyperparameters for the uses cases and architectures evaluated.

During this analysis, two differentiated phases have been performed. The first, where the best configuration of the network have been with respect to the neural network training hyperparameters, which included the learning rate and the batch size for both uses cases, whose results can be seen in Table 65. Additionally, for the infrastructure recognition the image size used to feed the neural network have been also analyzed, opting for using crops of 512x512. Furthermore, for YOLO architecture a study of its model size and the optimizer have been also performed, whose results, which led to use the medium model with stochastic gradient descent.

The second phase used this optimized hyperparameters to perform the long trainings to obtain the best performance of the architectures and datasets. Additionally, the use of the RGB weights to transfer the learning for SAR imagery is analyzed. Table 66 shows an overview of these sets of trainings with the best results for each of the dataset and architecture.

The employed datasets have shown the capabilities of the developed architectures for the use cases in military scenarios. In the case of the infrastructure recognition, the fusion of EO plus SAR imagery have resulted in a better performance in conditions of low visibility, such as foggy or smoky situations or where the clouds occlude the visible spectrum. In the same way, for the boat detection, the SAR imagery has proven that it could be a better spectrum than RGB-only cameras specially in off-shore situations, were the contrast between the see and the vessels is higher, and at night or low-vision circumstances.

Analyzing the datasets used, the SpaceNet 6 and HRSID have been the main datasets to reach the actual conclusions. The first, have allowed us with the more similar data respect to what it could be collected from a UAV with the ideal characteristics for this project. It provided with RGB and SAR images that were spatial-temporal synchronized, which allowed us to generate and test several fusion pipelines and with acceptable resolution images, that compared to TerraSAR-X, allowed us to analyze in the best possible way the performance of the architecture.

With respect HRSID and SSDD, we have a similar behavior that in the previous cases as the resolution and quantity is lower for the second. Nonetheless, the combination of the Airbus Ship Detection dataset and HRSID have shown that a first training in the visible spectrum could lead to a faster SAR training due to the transferring of knowledge, leading to having a EO and SAR system ready in a shorter time.

Use Case	Model	Dataset	Image Mode	Iterations	P (%)	R (%)	F1 (%)
1	UNET	SpaceNet	RGB	30000	0,873	0,638	0,737
1	UNET	SpaceNet	RGB-NIR	30000	0,862	0,638	0,733
1	UNET	INRIA	RGB	30000	0,804	0,852	0,804
1	UNET	INRIA	RGB+RGB	30000 / 30000	0,853	0,770	0,809
1	UNET	SpaceNet	SAR	30000	0,728	0,484	0,582
1	UNET	SpaceNet	RGB+SAR	30000 / 30000	0,736	0,499	0,595
1	UNET	Cloudy SpaceNet	RGB	30000	0,831	0,468	0,599
1	UNET	Cloudy SpaceNet	HSV-SPAN	30000	0,812	0,566	0,667
1	UNET	Cloudy SpaceNet	MEAN-PCA	30000	0,781	0,552	0,647
1	UNET	Cloudy SpaceNet	APPEND-PCA	30000	0,844	0,604	0,704
1	UNET	TerraSAR-X	SAR	30000	0,451	0,186	0,263
1	UNET	TerraSAR-X	RGB+SAR	30000 / 30000	0,434	0,218	0,290
1	UNET	Airbus	RGB	100000	0,906	0,888	0,894
2	YOLO	Airbus	RGB	100000	0,846	0,827	0,836
2	UNET	HRSID	SAR	60000	0,861	0,881	0,863
2	UNET	HRSID	RGB + SAR	100000 / 60000	0,875	0,875	0,866
2	YOLO	HRSID	SAR	60000	0,905	0,837	0,87
2	YOLO	HRSID	RGB + SAR	100000 / 60000	0,919	0,815	0,864
2	UNET	SSDD	SAR	60000	0,875	0,887	0,87
2	UNET	SSDD	RGB + SAR	100000 / 60000	0,919	0,815	0,864
2	YOLO	SSDD	SAR	60000	0,954	0,892	0,922
2	YOLO	SSDD	RGB + SAR	100000 / 60000	0,961	0,902	0,93
2	UNET	TerraSAR-X	SAR	60000	0,773	0,853	0,708
2	UNET	TerraSAR-X	RGB + SAR	100000 / 60000	0,891	0,841	0,782
2	YOLO	TerraSAR-X	SAR	60000	0,651	0,576	0,611
2	YOLO	TerraSAR-X	RGB + SAR	100000 / 60000	0,762	0,503	0,606

Table 7 Outputs of ARTINDET SC3: Results for UC1 and UC2.

The fusion pipelines implemented have proved promising results in the recognition of infrastructure, that was the only use case with an available dataset to evaluate them, that can be extrapolated to the ship detection. From the analysis of these methods, those which do not combine different data sources in the same channels lead to better results as the neural network is able to select the best data type depending on the situation. In general, EO imagery leverages a better performance due to the images having less noise, but it can be occluded by clouds, fog or smoke which SAR sensors can see through.

Regarding the architectures for the detection of ships, the semantic segmentation based neural network, UNet, have shown to be the most suitable choice for the task, especially if tracking is a later requirement. The main advantage with respect to YOLO is it the capability of returning the rotation of the vessels.

- **Next steps to be carried out for A-DRIT systems in the military domain.** A roadmap in the scope of DRIT systems in the military domain has been defined based on the work performed in ARTINDET SC 1, SC2, and SC3. This proposal covers the areas that require a higher effort in a short and mid-term perspective by analysing the business, data, application and technology gaps in the ambit defined by this project. The result includes a budget proposal, and the funding sources available to realize the corresponding projects. The summary of challenges and aims included in the roadmap are summarized in Table 8.

Challenge	Aim
Dataset definition and classification in the military domain	Methodology to select relevant visual features for military datasets
	Real military dataset compliance
	Simulated and synthetic military data generation
Applying supervised AI for target detection and tracking in military scenarios	Data fusion
	Urban
	Maritime
	Tracking
	AI deployment in embedded systems for real-time inference in different operational scenarios
Real unmanned systems with A-DRIT capabilities	Field tests in the military scope
	RPAS operative missions

Table 8 Outputs of ARTINDET SC2: New challenges and aims.

6. Acronyms

- **A-DRIT** - Automatic Detection, Recognition, Identification and Tracking
- **AP** - Average Precision
- **AI** - Artificial Intelligence
- **DRIT** - Detection, Recognition, Identification and Tracking
- **EO** - Electro-Optical
- **EU** - European Union
- **F1-Score** - Harmonic mean of precision and recall
- **HD** - High Definition
- **IR** – Infrared
- **mAP** - Mean Average Precision
- **SC** - Specific Contract
- **SSD** - Single Stage Detector
- **TRL** - Technology Readiness Level
- **UAV** - Unmanned Air Vehicle
- **UC** - Use Case
- **WFOV** - Wide Field of View