INCEEUROPE

CidB
Centre d'information sur le Bruit

QUIET DRONES
Second International e-Symposium
on
UAV/UAS Noise
27th to 30th June 2022

Deeplomatics: A deep-learning based multimodal approach for aerial drone detection and localization

Éric Bavu, LMSSC, Cnam Paris, HESAM Université, France
Hadrien Pujol, LMSSC, Cnam Paris, HESAM Université, France
Alexandre Garcia, LMSSC, Cnam Paris, HESAM Université, France
Christophe Langrenne, LMSSC, Cnam Paris, HESAM Université, France
Sébastien Hengy, French-German Research Institute of Saint-Louis, France
Oussama Rassy, French-German Research Institute of Saint-Louis, France
Nicolas Thome, CEDRIC, Cnam Paris, HESAM Université, France
Yannis Karmim, CEDRIC, Cnam Paris, HESAM Université, France
Stéphane Schertzer, French-German Research Institute of Saint-Louis, France
Alexis Matwyschuk, French-German Research Institute of Saint-Louis, France

Summary

Protection against illicit drone intrusions is a matter of great concern. The relative stealthy nature of UAVs makes their detection difficult. To address this issue, the Deeplomatics project provides a multimodal and modular approach, which combines the advantages of different systems, while adapting to various topologies of the areas to be secured. The originality lies in the fact that acoustic and optronic devices feed independent AI to simultaneously localize and identify the targets using both spatial audio and visual signatures.

Several microphone arrays are deployed on the area to be protected. Within its coverage area (about 15 hectares), each microphone array simultaneously localizes and identifies flying drones using a deep learning approach based on the BeamLearning network. Each array is attached to a local AI which processes spatial audio measurements in realtime (40 estimations per second), independently to the other units of the surveillance network.

A data fusion system refines the estimates provided by each of the AI-enhanced microphone arrays. This detected position is shared in real-time with an optronic system. Once this system has hooked its target, a Deep Learning tracking algorithm is used to allow an autonomous visual tracking and identification.

The optronic system is composed of various cameras (visible, thermal, and active imaging) mounted on a servo-turret. The active imaging system can capture scenes up to 1 km, and only captures objects within a given distance, which naturally excludes foreground and background from the image, and enhances the capabilities of computer vision.

The Deeplomatics project combines benefits from acoustics and optronics to ensure real-time localization and identification of drones, with a high precision (less than 7° of absolute 3D error, more than 90 % detection accuracy). The modular approach also allows to consider in the long term the addition of new capture systems such as electromagnetic radars.

1. Introduction

The illegal or hostile use of aerial drones is an emerging threat, which is only partially addressed by current ground or airborne anti-intrusion systems. The techniques required to identify moving targets with weak acoustic and visual signatures, and locating them for predictive trajectory tracking, represent more than ever a scientific and technical challenge. There are many applications related to defence in the context of securing sites, but also for locating targets thanks to compact and portable modules, which could complete the equipment of the 21st century soldier. They have many applications related to civil security (surveillance and security of critical energy access infrastructures, fight against industrial espionage, or security of demonstrations). These techniques are also of interest for civil applications in monitoring or controlling noise pollution caused by road or air vehicles, and for ecosystem monitoring applications (inventory and monitoring of animal species to protect biodiversity).

The DEEPLOMATICS project is aiming to achieve a scientific and technological leap forward to optimize multimodal detection and UAV threat tracking. We propose to integrate in an original way to the sensors of a surveillance network a set of independent artificial intelligences, specifically trained to respond to the tasks of real-time dynamic localization and target recognition. The majority of the project's tasks are based on a knowledge base acquired by the DEEPLOMATICS project partners in projects related to artificial intelligence for image recognition, acoustic source localization using Deep Learning, sound source recognition, but also in the development of sensors and specialized microphone arrays for the localization and imaging of acoustic sources, as well as in sniper detection projects, or acoustic beacons for helicopter detection. The DEEPLOMATICS project also involves an active imaging optronic system that has been adapted to automatic UAV identification and tracking using a real-time deep detector to perform drone recognition.

2. Multimodal sensors, Deep-Learning, and data fusion for UAV tracking and identification

2.1 Global system description

This interdisciplinary project uses advanced Deep Learning techniques, using the raw acoustic date measured by compact microphone arrays distributed over the site to be monitored,

complemented by an active imaging optronic system, which feeds an independent artificial intelligence for a computer vision task (see Figure 1.)

We believe that this modular surveillance network organization allows to adapt the sensor topology to the diversity of sites to be protected. The objective is to take advantage of the convergence of data-sciences, acoustics and optronic signal processing. When multiple acoustic and optronic systems are deployed in a fixed or reconfigurable manner at a site (urban or not) to locate a weak signature moving target, the first challenge is the real-time tracking of the moving target in a potentially noisy environment, and the orientation of the optronic systems towards the target.

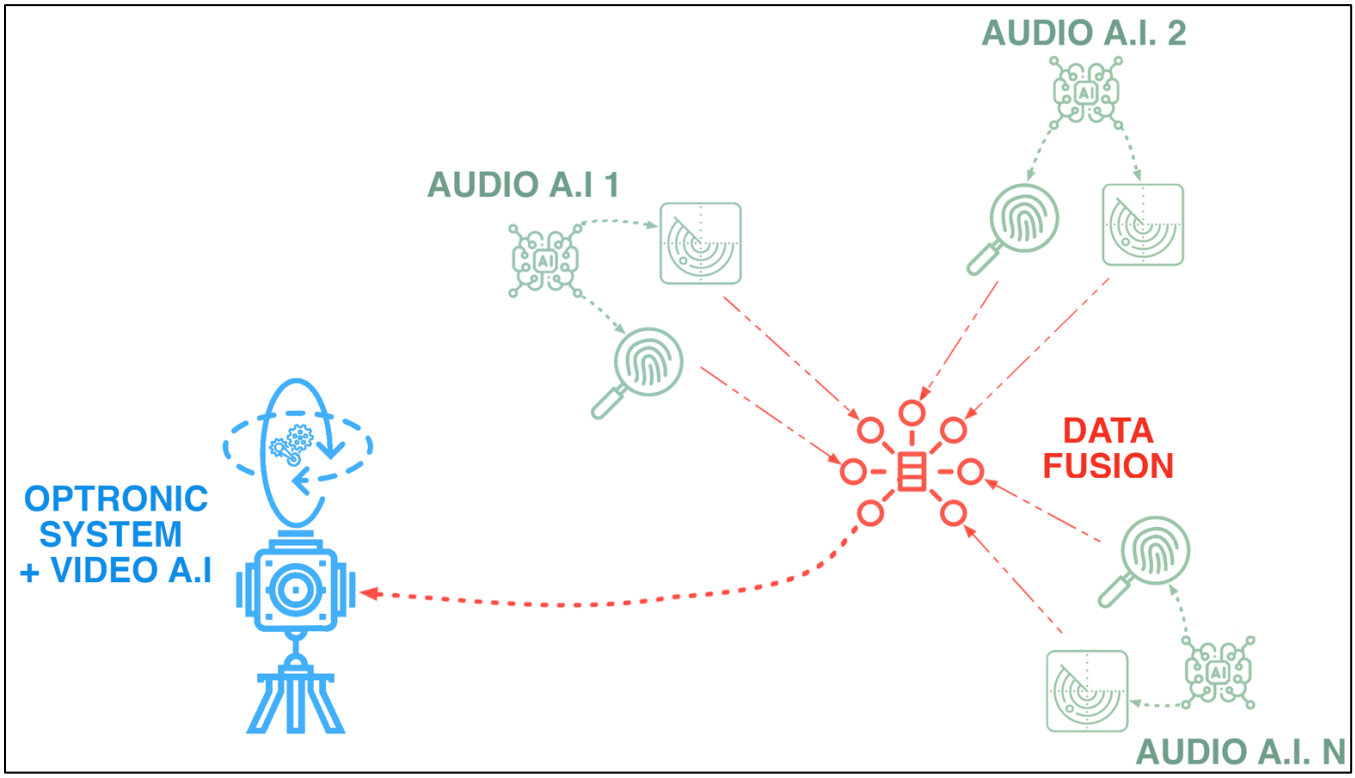

Figure 1: Multimodal detection and tracking using a set of N (3 depicted) A.I-enhanced microphone arrays, an optronic system feeding a realtime video drone detection A.I. The data fusion system refines the estimates provided by each of the AI-enhanced microphone arrays. This detected position is shared in real-time with an optronic system.

2.2 Acoustic surveillance network using A.I. units

For this purpose, the DEEPLOMATICS surveillance network is partly based on the use of a set of independent transportable broadband compact microphone arrays. The overall surveillance range using the audio modality is therefore only dependent on the number of AI-enhance microphone arrays in the surveillance network. The miniaturization of these microphone arrays is obtained thanks to the use of digital MEMS microphones. Their main advantages are their compactness, their adaptability, and their low cost. These microphone arrays, equipped with independent compact deep learning processors (see Figure 2), provide a solution adapted to the diversity of sites to be protected by recognizing the flying UAV while accurately identifying its position. The acoustic localization and recognition system will be further detailed in section 3 of the present paper.

Figure 2: AI-enhanced MEMS microphone array used in the project, with a compact, low-power AI processor (bottom right).

2.3 Video tracking

To confirm the presence of a UAV on the area to be protected, it is important to complete the information transmitted by the AI-enhance microphone arrays, which can sometimes generate false alarms, especially when many sound sources are present in the vicinity of the microphone array. Indeed, the trained acoustic deep learning networks allow a substantial screening of the detected and localized sound sources, but can generate false positive detections. For that purpose, an optronic system is also deployed in the area. This optical system is mounted on a motorized steerable turret stand and its orientation can be controlled by the data fusion application. In contrast to the microphone arrays, cameras have a much narrower solid angle of observation, but have the strong advantage of having a maximum range of 1 km, which can allow the video tracking of a non-cooperative UAV (see Figure 3). The optronic system is composed of various cameras (visible, thermal, and active imaging) mounted on a servo-turret.

The active imaging system can capture objects within a given selected distance, which naturally excludes foreground and background from the image [1,2], and can enhance the capabilities of computer vision . For example, when the drone blends into the background with the visible camera, active imaging can isolate the UAV by visually eliminating the background on the image. The parameters of these imaging systems are controlled by the fusion of information provided in real time by the AIs of each microphonic arrays placed on site.

In the Deeplomatics project, the images provided by the optronic systems are processed in real time to detect and track a drone present in the field of view of the optronic system. This task refers to the field of computer vision detection, which will be detailed in section 3, a task dominated today by deep neural network algorithms with convolution filters that perform by extracting visual features from the data. We decided to choose the YOLO [3,4,5] model as the final model. Its very fast inference time are perfectly with the constraint on the detection time per

frame was imposed so that the camera has time to adjust and track the drone. It was therefore necessary to have the fastest possible model.

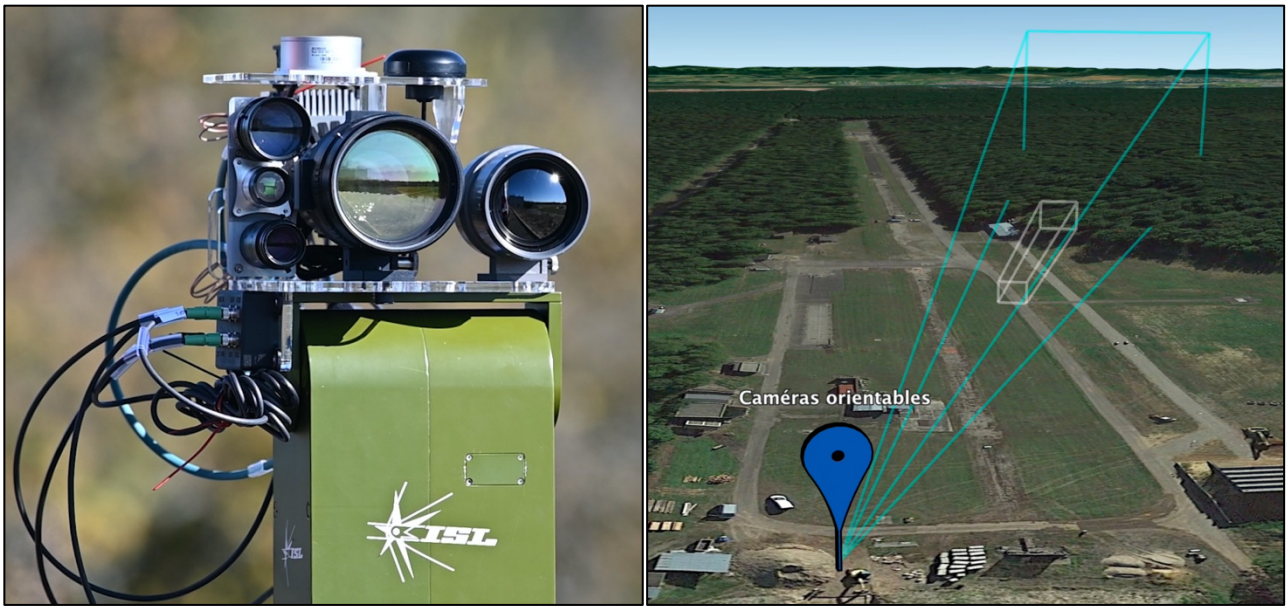

Figure 3: Left: The optronic system composed of various cameras (visible, thermal, and active imaging) mounted on a servo-turret. Right: On-site field of view of the various cameras used in the optronic system. Cyan: visible field of view. White: active imaging field of view. The active imaging range is selectable and controlled by the fusion unit which processes the inferred UAV positions transmitted in real time by the acoustic monitoring nodes.

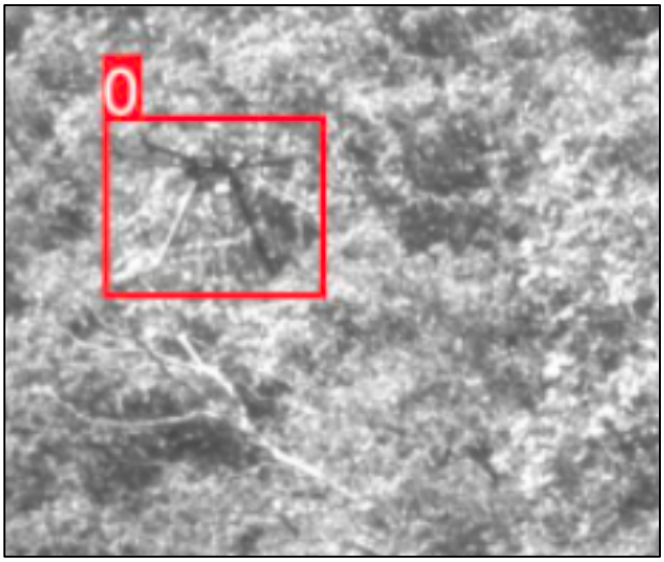

Figure 4: Exemple of a detection of a flying drone on a textured background using the trained YOLO v5 network using in-house dataset constituted during the project.

2.4 Data fusion

The data fusion application developed in the DEEPLOMATICS project must allow the analysis of data from different types of sensors deployed on the area to be monitored. Various types of sensors must be able to transmit information to the data fusion, including acoustic and optical sensors at this point. In future developments, the fusion should also be able to integrate

information from other types of sensors, including for example Radar, Lidar, and electromagnetic sensors.

Additionally, when monitoring a large area, the number of connected sensors can be large, so it is mandatory to establish a simplified information exchange, limiting the bandwidth used for communication. This information must then be processed quickly to locate the source with sufficient accuracy to be visible in the camera's field of view. To meet these constraints, the communication protocol between the sensors and the fusion application was defined based on the National Marine Electronics Association (NMEA) protocol, which was adapted to define "proprietary" messages. Using this data exchange protocol, the data fusion application manages the metadata transmitted by the different sensors present in the area to be monitored. The standard scenario consists in deploying several microphone arrays around the area to be protected in order to detect intrusions in the area. When a threat is detected, the data fusion allows to estimate its geographical position (latitude, longitude, altitude) and transmits this information to a camera which undertakes a second phase of detection/identification of the threat. In case of confirmed intrusion, the camera starts an independent tracking of the target and transmits information about its orientation to the data fusion. This information is then used to display the camera's field of view and its orientation on a map and to verify that the acoustic and optical data are consistent (see Figure 5).

In order to improve the tracking performance of a drone entering a sensitive area, a particle filter process is applied at various stages of the data fusion process. Particle filtering is a Bayesian recursive filtering method using discrete "particles" to approximate the posterior distribution of the system state. This filter has the advantage of being efficient whatever the distribution of the input data, but has a high computational cost because it requires a large number of "particles", i.e. samples, representative of the data distribution [6].

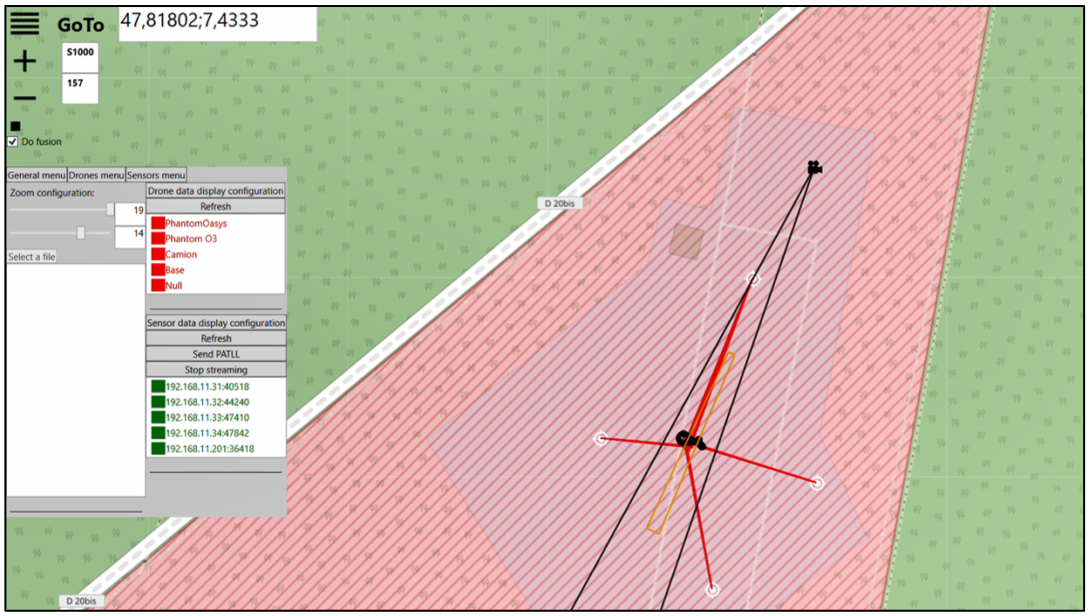

Figure 5: Human Machine Interface (HMI) of the fusion server integrating the position of four microphone arrays (white dots) and the associated estimated direction of arrival (red line), the position of the camera (black camera icon), its orientation and the associated field of view (black and orange lines). The fusion of the estimated UAV position provided by the 4 acoustic AIs allows to control the orientation of the visible and active imaging system to realize an automated video tracking of the drone.

3. Acoustic localization and detection using Deep Learning

In the DEEPLOMATICS project, each microphone array is attached to a deep neural network, trained for source localization and sound signature identification tasks. The neural network is a variant of the *BeamLearning* architecture [7] that we previously published for sound source localization. This variant of the network, *Beamlearning-ID,* has been specifically designed to simultaneously perform the recognition and localization tasks in real time [8]. The specialized AIs have been trained on a multi-channel dataset of acoustic signals from small UAVs in flight, under realistic conditions. These data acquisitions are augmented by a 3D spatializer. This augmentation will allow the neural network to respond as efficiently as possible to the localization and source identification tasks that will be performed simultaneously by the AI modules at the output of the compact microphone arrays.

3.1 Dataset: live measurements and higher order ambisonics 3D synthesis

A multichannel dataset of multichannel audio data was built throughout the Deeplomatics project to train the *Beamlearning-ID* network for drone localization and recognition. The acoustic signals recorded by the microphone arrays intrinsically convey information on the position of the acoustic source and its nature. The objective of the developed *BeamLearning-ID* deep network is to retrieve this information through supervised learning. Supervised learning requires a priori knowledge of this information. The audio data must therefore be annotated with the position and nature of the drone in flight.

The entire acoustic dataset is heavily annotated. To achieve this tedious task, a semi-automated process has been developed during the Deeplomatics project. During the flight of the drones, a GPS-RTK beacon is mounted on the drone and allows to know the position of the drone in real time. In parallel, several ambisonic microphone arrays record the 3D sound scene. The GPS and acoustic data are then synchronized. Moreover, a 3D spatialization step of the sound scene can be implemented if the antenna used to record the sound scene on site is different from the one used for the inference using the *BeamLearning-ID* network. This 3D spatialization process has been developed to produce an automated annotation of the multichannel audio (with labels denoting the drone model, and its 3D position synchronized with audio data) [8].

During the DEEPLOMATICS project, various measurement campaigns have allowed us to accumulate more than 34 hours of usable data of UAV flying data (simultaneous measurements of multichannel audio using high-order ambisonic microphone arrays and georeferenced position using a high precision realtime kinematics (RTK) GPS carried by the flying drones).

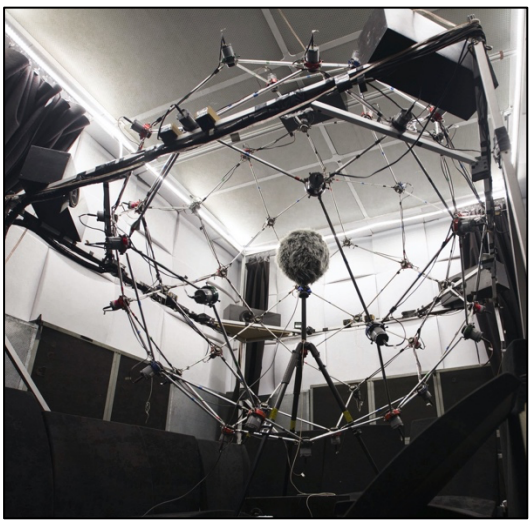

Figure 6: Higher order ambisonics spatializer used in the training process and dataset augmentation.

A large and realistic database allows deep neural networks to extract hidden patterns in the observation data. The size of the dataset is obviously not the whole story. For the deep neural network to be effective, it is necessary to build a dataset with a large variability of data. This is the reason why computer giants now have neural networks at their disposal that can exceed human capabilities in the field of image recognition. Image recognition researchers are now looking for ways to generate realistic synthetic images to train neural networks where the data sets are not yet large enough. We have, for localization or acoustic recognition tasks, access to a tool that allows to lift this lock and to generate simulated, augmented, or modified databases, while respecting the realism and the physical validity of the 3D pressure field.

The LMSSC has developed in the last few years a device that will allow to offer to localization and identification techniques by Deep Learning a flexibility and a realism not reached until now. Two tools developed and validated at the LMSSC are at our disposal [9-11]. The first device allows the spatialized capture of the sound environment, used in the measurement campaigns, and the second allows the restitution of the three-dimensional field (see Figures 6 and 7). These two devices allow us to render the 3D pressure fields of drones in flight on compact microphone arrays to train their individual artificial intelligence, even if the specific microphone arrays were not used for the on field experiments. One of the major advantages of this process is that we will also be able to "augment" the data captured during the measurement campaigns, by superimposing the 3D field of a large number of noisy environments, corresponding to potential locations for the installation of smart compact antennas (see Figure 7). These environments are also recorded by HOA ambisonic microphonic arrays.

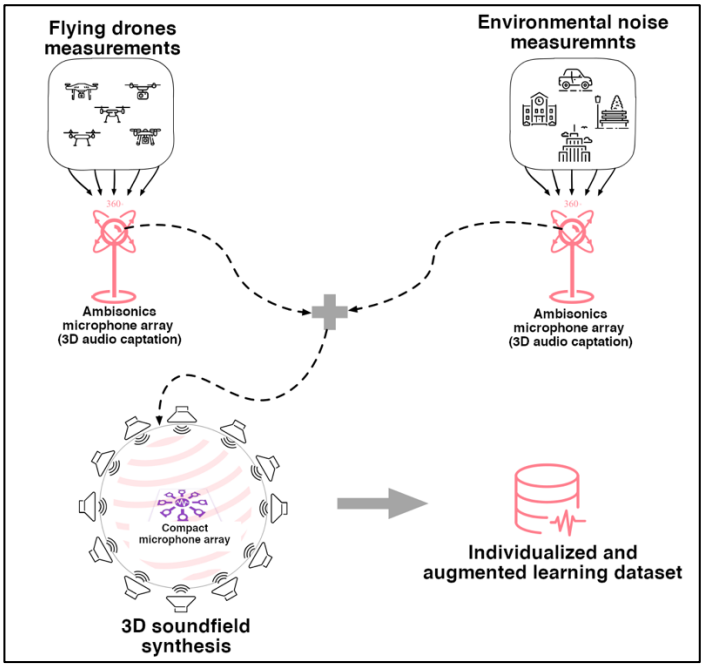

Figure 7: Schematic of the data augmentation strategy using the 3D spatializer for building an individualized audio learning dataset on a compact microphone: example of spatially noisy environment additions.

The flexibility of the ambisonic encoding operated by these two devices also allows us to modify the three-dimensional sound scene (e.g., modify the trajectory of the drone by rotation or dilation in the ambisonic domain / vary the signal-to-noise ratio and modify the spatial profiles of the ambient noise / etc ...).

The reproducibility of this physical synthesis of 3D acoustic fields will allow us to specifically train neural networks on different compact microphonic arrays, even other geometries than those used in the project. These AIs are trained to overcome or exploit the specificities of the environment in which the microphone will be installed, while implicitly performing a self-calibration of the several microphones included in the array [12]. This original approach provides Deep Learning for acoustics with the necessary variability to achieve abstraction and generalization capabilities that cannot be achieved by approaches based on array or environment models.

3.2 Beamlearning-ID deep neural network

Figure 8 shows the global architecture of *BeamLearning-ID* neural network developped specifically for this project. For more details on the underlying *BeamLearning* architecture, please refer to [7]. The BeamLearning-ID network is divided into blocks. The first block represents the raw input multichannel audio data, corresponding to the microphonic signals measured by the microphone array. The second block corresponds to a succession of several filter banks that allow to project the data into representative subspaces for the localization problem, thanks to residual subnetwork of atrous convolution kernels. The two parallel blocks in the third position are used to compute a pseudo-energy of the output channels of this succession of learnable filter banks, respectively for the localization and for the acoustic signature recognition tasks. Finally, the last two blocks allow to exploit these pseudo-energies, in order to deduce either the 3D angular position of the drone, or the type of drone having emitted the pressure field captured by the microphone array. The regression and classification approaches for source location have been compared by Tang *et al* [20]. In this project, the UAV angular localization problem, a regression approach is used. the source location will be given from a regression approach. Unlike the position of the source, the type of drone cannot be considered as part of a continuum. We therefore use a classification approach for the sound signature task. In our case, 6 classes are considered: one for the absence of drones and five for different drones used in this study (see Figure 9).

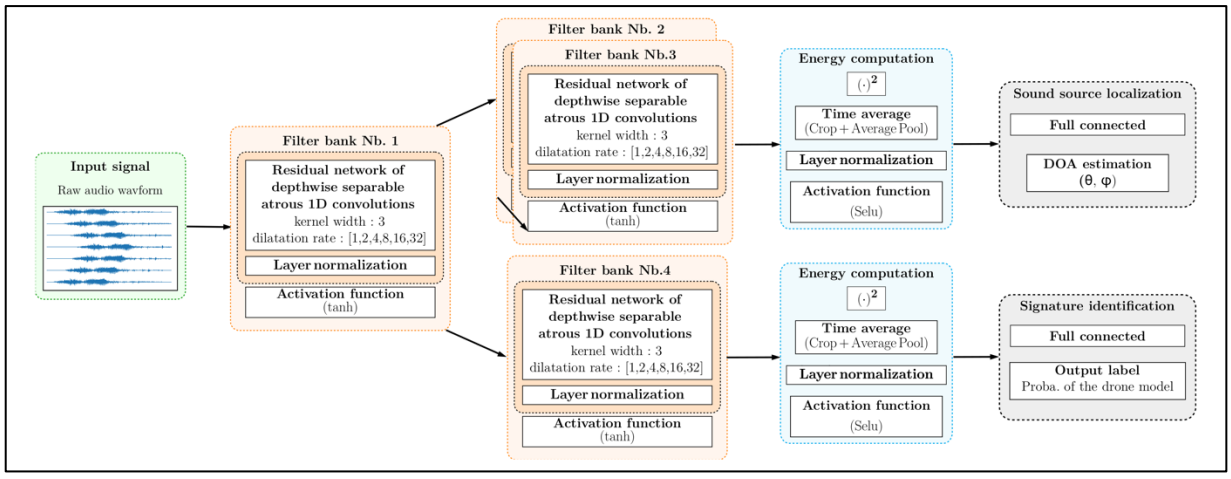

Figure 8: General architecture of the Beamlearning-ID network developed for the Deeplomatics project, consisting of two branches, one for drone recognition (bottom), one for realtime 3D localization (right)

Some deep learning architectures in the literature exploit pre-processed signals as input data, for example by using either the covariance of the signals, or their spectral representation, or the

information contained in the modulus, or/and the phase [14-20]. We propose, on contrary, to use raw temporal signals. The different convolutions used to process the data are precisely intended to project the temporal data into a representational space that is most appropriate to the problem at hand. Thus, we do not a priori constrain the data by pre-processing them. This approach, commonly called "Joint Feature Learning", represents an increasingly important area of research for Deep Learning applications in acoustics, and is since an a priori priori choice of representation for the input data can potentially omit features that the neural network could extract by itself. Moreover, thanks to this approach without pre-processing, the inference latencies are minimized and it is possible to maintain a real-time data processing approach.

Figure 9: Drones used for the flying drones dataset. From left to right: S1000, Phantom, Mavic pro 2, Mavic air, Spark.

3.3 Example of the localization and recognition performances for a single microphone array on a test flight

In order to illustrate the performance of the *Beamlearning-ID* trained network, this section presents the results of position and identification estimation obtained from a recording made by an AI-enhance microphone array (see Figure 2) during the June 2021 measurement campaign of the Deeplomatics project (data not used for the training process).

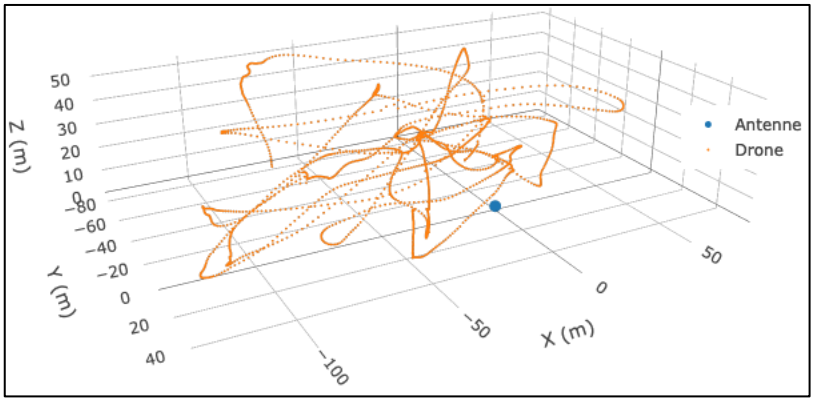

Figure 9: Relative position of the flying drone during the testing flight. Those positions are obtained using the mounted RTK-GPS beacon. Each point corresponds to a georeferenced position, sampled each 200 ms.

The drone used for this flight is an S1000 (see Figure 9). The flight lasted 7 min, which corresponds to 19734 consecutive estimates of drone positions and model identification (40 estimations/second). A 3D plot of the actual positions obtained using the RTK-GPS beacon mounted on the flying drone. Those positions are plotted on Figure 10 in a reference system centered on the microphone array, where the x axis points towards the north direction.

The way we designed the *Beamlearning-ID* deep neural network as well as its training process allows us not only to provide an angular position estimate at the output of the network, but also a confidence index noted r, which allows us to refine the estimated positions and to naturally filter the sound sources present in the environment of the microphonic array which are not flying drones. Figure 10 illustrates the statistical analysis for the angular localization performances for this flight, and Figure 11 illustrates the statistical analysis for the drone recognition task that is handled concurrently by the *Beamlearning-ID* network.

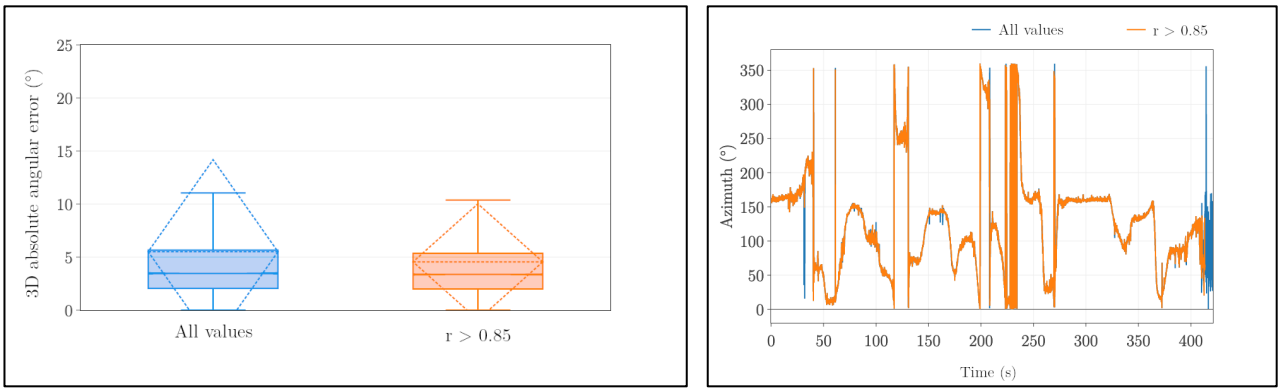

Figure 10: Left : Boxplot analysis of the 3D absolute angular error on the estimated position during the testing flight, without filtering data with the confidence index (blue), or with the use of the confidence index by only keeping the estimations that correspond to r > 0.85 (orange). Right : corresponding azimuthal estimations during the flight.

The analysis of figure 10 allows us to observe that the obtained 3D absolute angular errors are satisfactory during the whole flight, with a median of less than 4° (with or without the use of the confidence index as a filter). Using the confidence index to reject estimates due to non-UAV noise sources improves the results, with the mean 3D localization error improving by 16%, from 5.5° to 4.6°. On the other hand, the median varies only slightly, from 3.5° to 3.4°, which means that the confidence index has automatically removed outlier angular estimates due to auxiliary noise sources. This interpretation is confirmed by the estimated azimuthal trajectory plot on the right of Figure 10, especially at seconds 40 and 430: the estimates that are rejected are indeed estimates that are outliers with respect to the UAV trajectory.

The drone recognition functionality can also be evaluated. Figure 11 shows a histogram of the 19734 consecutive classifications obtained by the trained *Beamlearning-ID* network during the test flight presented above. The true-class inference rate is 76% for the raw data (in blue). On the other hand, after applying the r confidence criterion (in orange), the true-class inference rate is of 78%. The Deeplomatics project aims at protecting sites from drone overflights. Even if the recognized drone is not the right one, it is important that it is still recognized as a drone. The rate of non-detection of a drone observed in Figure 11 is 3% without using the confidence criterion and improves to 1% of non-detection of a drone on this flight. This observation confirms the effectiveness of the trained recognition system based on *Beamlearning-ID* archicture.

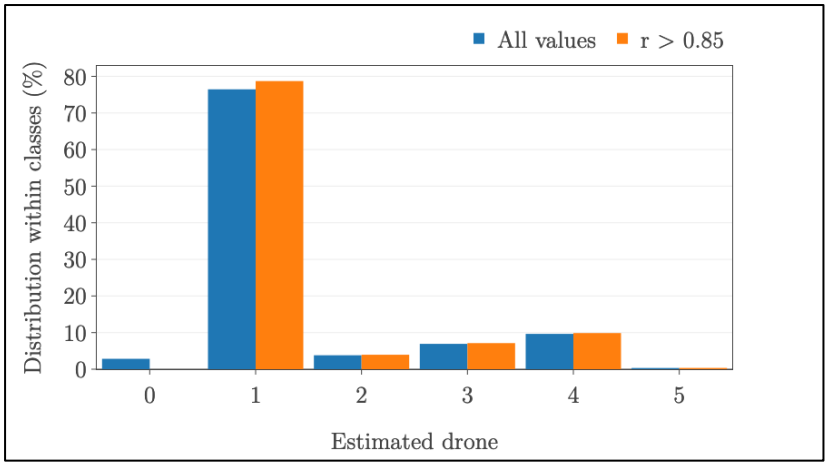

Figure 11: Drone recognition performances for the testing flight (total of 19734 estimations for 7 seconds of flight). The drone classes are 0 (no drone), 1 (S1000), 2 (Phantom), 3 (Mavic Pro), 4 (Mavic Air), 5 (Spark). The recognition histograms are shown without filtering data with the confidence index (blue), or with the use of the confidence index by only keeping the estimations that correspond to r > 0.85 (orange).

Thanks to the deep learning approach developed during the DEEPLOMATICS project, it is therefore possible to detect, localize and recognize a drone intrusion using a single AI-enhanced microphone array in its coverage area. The main benefit of the proposed approach is to perform these three tasks simultaneously which allows to spare a significant amount of time during the estimation process. With this approach, it is actually feasible to perform these three tasks in real time on relatively light hardware architectures (see Figure 2).

4. Conclusions

All the technological bricks of the Deeplomatics project are now functional and interoperate in realtime. Each microphone array associated to its own Beamlearning-ID network allows to detect and localize a drone intrusion, at a rate of 40 estimations per second. The estimations of each microphone array are sent in realtime to the data fusion unit in order to refine the georeferenced position of the drone in flight and its identification. The analysis of the output data of the fusion unit shows that for all the flights tested, the position error obtained is on average 13 meters when the drone is in the middle of the acoustic antenna cluster, ensuring the presence of the threat in the camera's field of view when the camera is 200 meters away from the microphone array cluster. Further developments concerning the acoustic devices include the industrialization of custom microphonic arrays with custom AI processors, and the potential use of informed spatial filtering in order to improve the detection and localization range.

Acknowledgements

This work has been funded by the DGA/AID Grant No. ANR-18-ASTR-0008.

References

[1] Christnacher, F., Monnin, D., Laurenzis, M., Lutz, Y., & Matwyschuk, A. (2011). Imagerie active: la maturité des systèmes ouvre de vastes perspectives. *Photoniques*, (55), 44-51.

[2] F. Christnacher, S. Hengy, M. Laurenzis, A. Matwyschuk, P. Naz, S. Schertzer et G. Schmitt, «Optical and acoustical UAV detection,» *Proc. of SPIE Security + Defence 2016,* vol. 9988, 2016.
[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
[4] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
[5] Glenn Jocher et al. ultralytics/yolov5 : v3.1 - Bug Fixes and Performance Improvements. Version v3.1. Oct. 2020. doi : 10.5281/zenodo.4154370. url : https://doi.org/ 10.5281/zenodo.4154370.
[6] Orlande, H., et al., Kalman and Particle filters. METTI V-Thermal Measurements and Inverse Techniques, 2011.
[7] Pujol, H., Bavu, E., & Garcia, A. (2021). BeamLearning: an end-to-end Deep Learning approach for the angular localization of sound sources using raw multichannel acoustic pressure data. *The Journal of the Acoustical Society of America*, *149*(6), 4248-4263.
[8] Pujol, H., Bavu, E., Garcia, A., Langrenne C., Hengy S., Schertzer S., Matwyschuk A. (2022, April). Deeplomatics : Localisation et reconnaissance acoustique de drones . In *16ème Congrès Français d'Acoustique, CFA 2020*.
[9] P. Lecomte, Ambisonie d'ordre élevé en trois dimensions: captation, transformations et décodage adaptatifs de champs sonores, Thèse de doctorat: Paris, CNAM, 2016.
[10] P. Lecomte, P. A. Gauthier, C. Langrenne, A. Garcia et A. Berry, «On the use of a Lebedev grid for ambisonics,» *Audio Engineering Society Convention,* 2015.
[11] P. Lecomte, P. A. Gauthier, C. Langrenne, A. Berry et A. Garcia, «Cancellation of room reflections over an extended area using Ambisonics,» *The Journal of the Acoustical Society of America,* vol. 143(2), pp. 811-828, 2018.
[12] Bavu, E., Pujol, H., & Garcia, A. (2018, April). Antennes non calibrées, suivi métrologique et problemes inverses: une approche par Deep Learning. In *14ème Congrès Français d'Acoustique, CFA 2018*.
[13] Eric L Ferguson, Stefan B Williams, and Craig T Jin. Sound source localization in a multipath environment using convolutional neural networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages pp.2386–2390. IEEE, 2018.
[14] Weipeng He, Petr Motlicek, and Jean-Marc Odobez. Deep neural networks for multiple speaker detection and localization. In 2018 IEEE International Conference on Robotics and Automation (ICRA), pages 74–79. IEEE, 2018.
[15] Ryu Takeda and Kazunori Komatani. Sound source localization based on deep neural networks with directional activate function exploiting phase information. In IEEE International Conference on Acoustics, speech and Signal Processing (ICASSP), pages pp.405–409. IEEE, 2016.
[16] Fabio Vesperini, Paolo Vecchiotti, Emanuele Principi, Stefano Squartini, and Francesco Piazza. A neural network based algorithm for speaker localization in a multi-room en- vironment. In 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), pages 1–6. IEEE, 2016
[17] Sharath Adavanne, Archontis Politis, and Tuomas Virtanen. A Multi-room Reverberant Dataset for Sound Event Localization and Detection. In Submitted to Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), Munich, Germany, 2019.
[18] Soumitro Chakrabarty and Emanuel A.P. Habets. Multi-speaker DOA estimation using deep convolutional networks trained with noise signals. IEEE Journal of Selected Topics in Signal Processing, Vol. 13(No. 1) :pp.8–21, 2019.
[19] Zhenyu Tang, John D Kanu, Kevin Hogan, and Dinesh Manocha. Regression and classification for direction-of-arrival estimation with convolutional recurrent neural networks. arXiv preprint arXiv :1904.08452, 2019.