A Multimodal AI-Leveraged Counter-UAV Framework for Diverse Environments

Eleni Diamantidou, Antonios Lalas, Konstaninos Votis, Dimitrios Tzovaras

A Multimodal AI-leveraged Counter-UAV Framework for Diverse Environments*

Eleni Diamantidou[0000−0001−6763−054X], Antonios Lalas[0000−0002−5337−161X], Konstaninos Votis[0000−0001−6381−8326], and Dimitrios Tzovaras[0000−0001−6915−6722]

Centre for Research and Technology – Hellas (CERTH), Information Technologies Institute, 6th km Harilaou - Thermi,57001 Thessaloniki, Greece
{ediamantidou,lalas,kvotis,dimitrios.tzovaras}@iti.gr

Abstract. Unmanned Aerial Vehicles (UAVs) have become a major part of everyday life, as well as an emerging research field, by establishing their versatility in a variety of applications. Nevertheless, this rapid spread of UAVs reputation has provoked serious security issues that can probably affect homeland security. Defence communities have started to investigate large field-of-view sensor-based methods to enable various civil protection applications, including the detection and localisation of flying threat objects. Counter-UAV (c-UAV) detection challenges may be granted from a fusion of sensors to enhance the confidence of flying threats identification. The real-time monitoring of the environment is absolutely rigorous and demands accurate methods to detect promptly the occurrence of harmful conditions. Deep learning (DL) based techniques are capable of tackling the challenges that are associated with generic objects detection and explicitly UAV identification. In this paper, we present a novel multimodal DL methodology that combines data from individual unimodal approaches that are associated with UAV detection. Specifically, this work aims to identify and classify potential targets of UAVs based on fusion methods in two different cases of operational environments, i.e. rural and urban scenarios. A dedicated architecture is designed based on the development of deep neural networks (DNNs) frameworks that has been trained and validated employing real UAV flights scenarios. The proposed approach has achieved prominent detection accuracies over different background environments, exhibiting potential employment even in major defence applications.

Keywords: Unmanned Aerial Vehicles (UAVs) · Multimodal Deep Learning · UAV detection.

1 Introduction

Unmanned Aerial Vehicles (UAVs) have gained a considerable part in several technological applications. In recent years, many industries have involved UAVs

* This work was supported by the European Union's Horizon 2020 Research and Innovation Programme Advanced holistic Adverse Drone Detection, Identification and Neutralisation (ALADDIN) under Grant Agreement No. 740859

to produce solutions that improve time and resources consuming demands. Additionally, there has been a significant increase in UAVs training and practice, which has made the UAVs a friendly solution for the open public. A major part of applications that employ UAVs is associated with environment tracing such as vehicle tracking, natural disasters detection, reconnaissance of suspicious actions or cargo transportation [21]. However, there are a lot of restrictions regarding UAV flight plans and flight approvals. In addition, there is a possibility that a UAV may not have an authorised and appropriate flight plan [23]. According to many European countries authorities, UAVs should not fly over people, prisons, hospitals, government, military facilities, other sensitive areas or airports. Following that, counter-UAV (c-UAV) systems have introduced a new research field concerning national safety, since this novel technology is capable of causing security threats and harm civil protection. Due to their lightweight and small size, low vibration, low thermal cost and high power to weight ratio, the UAVs can conveniently fly at any environment including both urban and rural areas with the minimum level of intrusiveness on the environment [13]. Thus, finding an effective solution to accomplish accurate UAV detection is deemed necessary for security purposes, acquiring at the same time great attention from the research community. A variety of c-UAV solutions involve, among other, multi-sensory methods including vision, radar, acoustic or radio-frequency sensors.

A high-interest challenge relies on how efficiently a deep learning (DL) fusion model that utilises vision and radar schemes can identify the presence of UAVs near cities or in regions located in the outer parts of the cities. Since the different performance of the model may be critical for security purposes, the knowledge about the effects of the surrounding environment to the overall detection capabilities may pose a significant factor for the effectiveness of the final c-UAVs system. In this work, a multimodal DL based method is presented that employs thermal and 2D radar data, to adequately tackle aerial vehicle detection problems in cases where a flying threat is approaching an essential urban or rural infrastructure without permission. For this purpose, dedicated data captures have been conducted in different environments allowing for the training and comparison of topology-oriented models for accurate detection. Explicitly, this work aims to efficiently introduce a multimodal neural network-based algorithm for the UAV detection task in diverse environments, taking into account the different conditions that may be applied, and assess their impact in the final performance of the holistic system.

The rest of the paper is organised as follows. Section 2 provides an overview of recent studies regarding DL sensing methods for UAV detection systems. Section 3 is devoted to the detailed description of the UAV flights environments, together with the sensing modalities that can be utilised to address the UAV detection task. Also in this context, the recommended DL classification activities regarding the UAV detection task are proposed in Section 4. The experimental process as well as the corresponding evaluation procedure and results are provided in Section 5. Finally, Section 6 summarises the conclusions of this work.

2 Literature review

Many novel multimodal techniques that employ multiple sensors and data fusion have been proposed in an attempt to increase accuracy of detection. An indicative example is associated with DL technology which established very promising approaches concerning sensing topologies and vehicle detection. In the context of c-UAV applications, bearing-only and radar sensors had been fused to detect and localise UAVs [7]. The employment of these sensors resulted in false alarm (FA) discrimination. In addition, the track extraction time was reduced, which corresponds to the necessary time that the system requires to output decisions. Another interesting work [3] proposed a methodology of radar and vision fusion, aiming to improve vehicle detection and identification reliability. The authors presented a high-level fusion of radar and vision fusion targets. This system utilised a video system to validate the radar detections and enhance the overall accuracy. The evaluation of this method was performed in urban environments achieving great results. Vision and radar systems remain among the well-known approaches to detect, identify and localise flying threats. Each sensor exhibits advantages and disadvantages in different scenarios, which affect the final detection decisions owning that the field of view varies depending on the background of the associated sensor that tracks a target [19]. Moreover, sensors have limitations that are related to the environment such as weather conditions and external noise. A principal task is to join different sensing modalities to overcome or compensate for possible detection weaknesses. Multi-sensor information methodologies indicate their applicability on UAV detection since they are capable of providing considerable advantages over single-sensor information [6]. Explicitly, this work, even though limited to a specific environment only, has efficiently introduced a neural network-based algorithm for the UAV detection task, formulating a general information fusion framework that merges extracted features from multiple modalities.

3 Description of the Environment and sensing methodologies

In recent years, UAVs have shown immediate growth in everyday life, generating many security issues. The commercial usage of UAVs has been extended to a variety of areas, where UAVs can regularly fly at different backgrounds without being noticed. Taking into account that UAVs have also a small size and they move quickly, the detection challenge within diverse environment has been made intensively complicated.

3.1 Multi-sensory technology

Many available sensors can identify and localise flying threats. The majority of them refer to vision and radar systems. In our work, we take advantage of thermal imaging and 2D radar sensing capabilities. Both the 2D radar sensor

and the panoramic thermal camera capture the predefined area over a long range in a 360 degrees point of view. This allows the sensors to perceive all the states of the environment. The 2D radar sensor functionality is based on the emission of radio waves to determine the range, the azimuth or the velocity of targets. As many targets as possible exist, there will also be many 2D radar reflections [20]. However, since UAVs are small and tend to fly close to the ground, even the more specialised radar sensors can face difficulties to identify and classify them as UAV threats. Regarding the thermal sensors, they illustrate great potential for UAV detection by detecting their heat signature. Thermal cameras detect flying targets at specific azimuths and elevations by recognising different levels of infrared radiation [24]. According to this, these sensors have a great sensitivity to weather conditions such as high temperature and humidity. Subsequently, each sensor imposes its advantages and disadvantages in several scenarios making the single-sensor UAV detection a demanding task. Nevertheless, it is tolerable to obtain an effective solution, especially assuming that supplemental technologies are incorporated to ensure maximum coverage in any possible gaps.

3.2 UAV flight regions

As mentioned above, UAVs can effortlessly fly over individual areas without drawing attention. As a result, it is essential to implement dedicated algorithms that have the ability to learn to identify UAV threats in several environments. Intending to appropriately address this research challenge, it appears a necessity to gather plenty of UAV detection data that correspond in different environments. Besides, the majority of algorithms that involve predictive capabilities are associated with DL techniques, which certainly demand considerable amounts of data [18], to discover patterns between the potential targets [18]. For these reasons, various data collection sessions were fulfilled during the European project ALADDIN [1] to capture UAV flights under different circumstances. The varying backgrounds and the several types of aerial vehicles produced an explicit and valuable dataset for DL research activities. The first data gathering session took place in a rural area at the Air Traffic Laboratory for Advanced unmanned Systems (ATLAS) in Villacarrillo, Spain. The remaining data gathering sessions occurred at an urban area on Markopoulo premises in Athens, Greece. ATLAS, that Figure 1[a] represents, refers to a test flight centre located in an open field location, surrounded by trees. On the other hand, the Markopoulo Training Facility, presented in Figure 1[b] is settled just outside the city of Athens. The aforementioned regions were selected to accurately gather data regarding UAV flights at both rural and urban environments, thus enabling comparison of the environment variability.

To further elaborate on the environment analysis of the UAV flight recordings, there are some variations of major importance between a rural and an urban area. The most notable difference between a rural and an urban scene is associated to the external environmental noise. UAV tracking in regions near cities can be affected by FAs similar to car traffic or by buildings that can cover possible threats. On the other hand, countryside areas involve less FAs. Despite

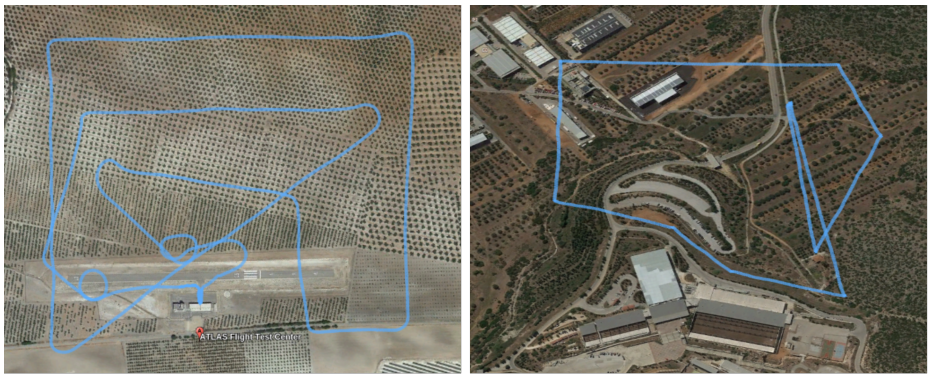

Fig. 1. UAV Test Flights. [a] The flight plan captured at ATLAS, was designed to simulate the intrusion of a protected area from the dark side of the hangars. [b] The flight plan captured at Markopoulo involved incoming drones from the dark side of the buildings of the protected area.

that, the UAVs can still hide, for instance at trees, during a flight operation, making the identification between UAVs and FAs a really difficult task. Our DL fusion model has the virtue to tackle the above challenges. The main goal is to achieve a reliable UAV detection at both rural and urban environments with maximum accuracy.

4 Deep learning technology for UAV detection

Various sensor-based methods utilise tracking trajectory techniques for flying vehicles to recognise their flight signature and decide if the vehicles are possible threats [5,14,16]. DL based methodologies have gained great attention in a wide range of vehicle detection over the past few years since they have shown successful results [12]. In a general sense, DL attempts to learn significant features from data, understand patterns and relations between them to obtain predictive power in new unseen data [9]. The Deep Neural Networks (DNNs) have the ability to learn high-level representations from data and recognise relations among them and the target variable to predict. The concepts of DL may undoubtedly further apply their findings in problem cases where the data are coming from multiple resources [15]. In the majority of world problems, the data arrive from multiple sources. The most interesting challenge in DL data fusion is to perform a joint representation of multimodal data [2]. A fundamental point is related to the way a DNN can detect patterns and associations between these data. Multimodal DL has to discover associations that link different modalities and join them. The predictive behaviour of a DL neural network derives from multi-layered and sometimes complicated structures [11]. Taking this into account, the findings of a Multi-layer Perceptron classifier are adopted [17].

4.1 Training Data

The proper formulation of a multi-input neural network architecture requires a joint data representation. A solution to this problem could be derived by the construction of a shared representation. As already noted, DNNs can learn significant relations and features from the input data [25]. Consequently, DNNs can be valuable feature extractors producing high-level representations from data. These high-level representations are in the form of large numerical feature maps that have been output from single DL feature extractors [8]. In the UAV detection problem, the input samples belong to two training classes: the UAV class and the FA class. Diving into the content of each training class, the FA class is characterised as a highly generic class, since FAs in a real-world scenario can be birds, trees, some clouds, sensor noise or even humans and cars referring to an urban environment. On the other hand, the UAV class is very precise, since it is only associated with several types of UAVs targets.

To efficiently combine thermal and 2D radar data to accomplish the task of UAV detection, two unimodal DNNs were utilised to generate a common representation at both input streams. The thermal camera and the 2D radar have a diverse perception of identifying targets in between them. The thermal camera features result in a three-dimensional signal considering that the information is extracted from images, whereas the 2D radar generates a two-dimensional signal. Therefore, each sensor ends with a unique type of detection. Following that, it is essential to discover an effective way to join multimodal information despite the different data meaning, shapes and types. In addition to that, handling data captured from different sensors leads to a necessity of data alignment. The input data refer to different capturing sessions and recordings using various types of UAVs under different time frames, spatial ranges, weather conditions, etc. allowing a wide diversity in the dataset. In addition, multimodal data require a matching process to properly align them in at least the temporal domain. Accordingly, an alignment method based on temporal adjustment was implemented. This process uses timestamps provided from each capture of each modality accordingly.

4.2 UAV detection algorithm architecture

The aim of the proposed fusion algorithm is related to the task of classification and identification to enhance the UAV detection accuracy as data from two sources are combined. Our DL fusion neural network attempts to identify the presence of flying targets in two major scenarios. The first scenario is assigned to UAVs that fly in the countryside. The second scenario refers to UAV flights that are operated near urban areas. Figure 2 represents the concept and the design of the recommended architecture of the DL data fusion framework.

The neural network architecture has been designed utilising two input streams: high-level representations from thermal images and 2D radar signatures. Combining different modalities for enhancing the detection accuracy seems an intuitively appealing task [6]. Since the thermal and the 2D radar features belong in

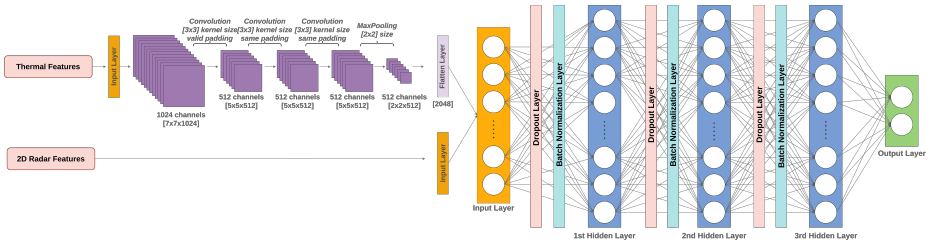

Fig. 2. Proposed architecture of fusion neural network for UAV classification. The neural network has two input streams that correspond to 2D radar data and thermal imaging data.

different dimensional space, they will have a different impact on the prediction output. The thermal data enable feature maps of 7x7x1024 dimension, while the 2D radar data outputs feature maps of 1664 dimension. It is noted that the thermal input is considerably larger than the 2D radar input which seemingly results in a possible misleading learning process. To overcome dimensionality diversities and avoid displeasing learning issues, a further process of thermal data was implemented. The main solution concept relates to convolutions. Employing convolutions aims to shrink the multidimensional input without losing possible meaningful information [10]. The implementation of the proposed methodology includes a concatenation layer to combine the two input streams. Each stream is related to a different input tensor. The output of the concatenation layer returns a single tensor that contains the concatenation of all inputs. Several additional parameters can affect the training procedure. Among the major neural network initialisation parameters are the number of hidden units, the number of hidden layers and the optimisation method that the algorithm will update and adjust itself. In particular, the number of hidden units depends on the size of the dataset. Very few hidden units may produce generalisation errors due to under-fitting effects. Too many hidden units may result in low training errors, but will make the training unnecessarily slow, and will lead in poor generalisation. Concerning the training process, the weights are initialised randomly, which means that the neural network is training without using weights of other pre-trained models. The Adam optimiser was adopted to adjust the DNN weights, while the learning rate was initialised equal to 0.001. Regarding the two-input streams neural network, it consists of one input layer that contains the joined thermal and 2D radar features, three fully connected hidden layers (dense layers) and one output layer. The input of the data fusion neural network is the concatenation of multiple unimodal features, as mentioned earlier. However, the output of the data fusion neural network is a binary problem classification result. Specifically, the output of our DNN is a probability according to the training classes: UAVs or FAs. Dropout layers have been employed to prevent overfitting during neural network learning [22]. Correspondingly, batch normalisation layers have been

also employed to normalise input data [4] since they are coming from multiple resources and processes.

5 Experimental Process and results

The data fusion concepts support two main purposes. The first purpose relates to detection accuracy enhancement while adding multiple modalities. The second and most meaningful purpose is relevant to missing modalities completion. Aiming to focus on a comprehensive examination about the fusion input data, two major experiments were conducted. According to this, the experiments utilised large amounts of data that represent individual cases of the problem to gain significant predictive power. These experiments aimed to identify which environmental setups have the merit of fusion concepts. For this reason, our neural network model architecture was trained on two diverse datasets to produce robust models that can identify UAVs in rural and urban areas. The two datasets involved high-level representations of thermal and 2D radar UAV detections recorded from the ATLAS flight centre and Markopoulo, referring to a rural and urban environment respectively.

The investigation on the classification behaviour and the evaluation performance of a proposed architecture using data with background diversities is certainly appealing. Powerful DL models need a suitable selection of the data that the neural network would train on and likewise the test data. The training set contains a known output and the model learns on this data to be generalised to other data later on. Assuming that, the train and validation processes would be applied to specific data during the learning process. In our experimental processes, both the fusion model with UAV detections recorded in the rural area and the fusion model related to the UAV flights near cities, employed training and validation samples involving an equal number of instances at both training classes. To verify the performance of our algorithms, the evaluation of the fusion models included UAV recordings with clear flight plans. Besides, to ensure the quality of the results, the UAV recordings that were selected for evaluation purposes, had as much as possible, similar flight plans. As a result, to properly test the fusion models and measure evaluation performance, two UAV recordings that follow the same notions of a flight plan were selected in both cases of the environmental setup. The main evaluation scenario refers to a UAV threat coming from far away. While the UAV is approaching the region of interest, it is hiding at protected areas and moving without being easily noticed. Both evaluation recordings were captured at daytime and the UAVs flew approaching a maximum 700m distance from the sensors. Figure 1[a] demonstrates the flight plan of a UAV in a rural environment. In the same way, Figure 1[b] involves the flight plan of a UAV in an urban environment.

For the evaluation process, the neural network was initialised with a set for evaluation samples to monitor the classification output based on ground truth labels, which in our case refers to the presence of a UAV. As a result, confusion matrices were utilised for examination of fusion model performance. A confusion

matrix is an excellent performance measurement for classification problems. Each row of the confusion matrix represents the instances in a predicted class, while each column represents the instances in an actual class. It is necessary to remark that test flights correspond to real UAV flights scenarios. Accordingly, possible differences in the number of detection in each class are expected. Regarding the validation of the fusion model associated with rural area UAV flights, the confusion matrix in 3[a] describes that from the 472 positive samples that represent UAV detections, 417 samples were correctly classified as UAV, and only 55 samples were classified mistakenly as FAs. Likewise, among 592 negative samples, 475 samples were correctly classified as FAs, and 117 samples were classified wrongly as UAVs. In the same notions, the fusion model validation associated with urban area UAV flights is performed in Figure 3[b]. The test flight captured

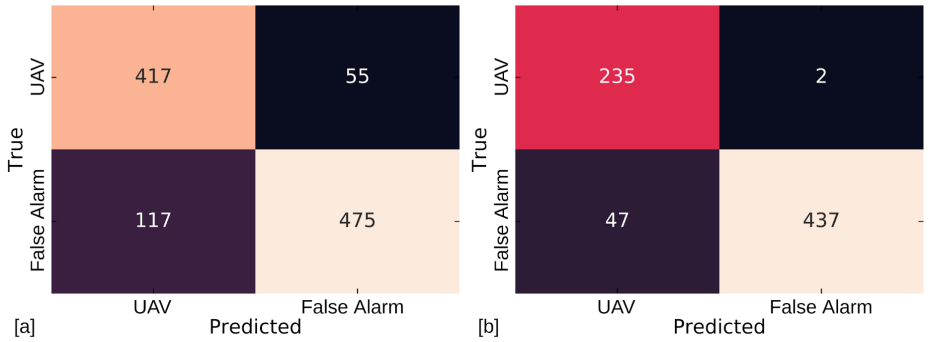

Fig. 3. Confusion matrices provide information about the validation performance of the multi-sensory model on target data points. Specifically, confusion matrix regarding the evaluation of the fusion model trained on data points recorded in [a] rural areas, and [b] urban areas.

in an urban environment involved 237 positive samples, where the largest part of detections was classified perfectly and only 2 samples were falsely classified as negatives. However, among 484 FA detections, 437 data points were accurately classified as FAs, and only 47 negative samples were inaccurately classified as UAV detections.

Accomplishing the task of fusion model evaluation, it is expected that the fusion model will succeed in the challenge of distinguishing UAVs from FAs in several cases. For quality estimation and efficient comparison between the two cases of the experiment, the precision, recall and the F1 score metrics were utilised since they are of the most prominent measures on a classification problem. As it was assumed, the two fusion models achieved exceptional performance at both the distinct environment training experiments. The evaluation metrics concerning the rural test flight feature samples are encouraging. Likewise, the results that are referring to the urban test flight sample have shown even more exceptional performance as described in Table 1 and Table 2.

Table 1. Experimental results that present the performance of UAV classification task at rural environments

	Precision	**Recall**	**F1 Score**
UAV	0.7808	0.8834	0.8290
False Alarm	0.8962	0.8023	0.8467

Table 2. Experimental results that present the performance of UAV classification task at urban environments

	Precision	**Recall**	**F1 Score**
UAV	0.83	0.99	0.91
False Alarm	0.99	0.90	0.95

Many deductions may be derived during the examination of the outcomes that associate to the above mentioned experiments. At a first though, it may be expected that the performance of the fusion model that was trained by handling samples recorded in the rural environment should be higher than the related fusion model performance that used UAV detections captured in an urban environment. In a rural area, the UAVs are an easily recognisable target since the background is clear free of additional environmental noise. Considering this, the evaluation metrics of the rural area fusion model are expected to be improved compared to the evaluation metrics of the urban area fusion model. However, a different result arises, where the urban case exhibits better evaluation metrics. The principal explanation to this result relies on that the dataset that indicates the UAV flights in the countryside includes FAs which consist of fully generic detections. Namely, there are few FAs in the countryside, which could describe small parts of clouds, trees or birds. Moreover, in some cases, FAs may have been introduced by capturing sensor noise. In a public region, however, there is an increased number of well-defined false warnings. In this particular case, FAs can refer to buildings, public transport, aeroplanes or any other external noise. A classification problem requires training classes that consist of explicit data that precisely represent the desirable target. According to this, our experiment has shown that the same multimodal neural network architecture can classify more convenient UAV detections in an urban environment compared to a rural environment. This is mainly explained due to the well-defined nature of the false warnings in urban cases in comparison to the rural environment, where the cause of FAs is more obscure. Bearing in mind that the UAV, as well as the FA detections, were expressed equally in both cases of experimentation, our fusion model achieved higher evaluation performance at UAV detection data that are associated with regions near cities.

6 Discussion and Recommendations

This research aims to understand in an enhanced way, the learning behaviour of UAV high-level representation detections in two different backgrounds. In this context, complex experiments with individual input vectors were conducted to

identify cases where the fusion model exhibited the merits of the multimodal learning concepts. The involved 2D radar and thermal sensors allowed for a global field of view which did not result in homogeneous data. Thus, the amount of all single-sensor feature maps varies, based on the limitations of each sensor. To overcome these boundaries, our proposed model architecture was based on robust DL architectures. As a result, we have successfully demonstrated the effectiveness of the multimodal data learning and appropriately established our efficient fusion model, which is suitable for efficient UAV detection in several environments. Moreover, an enhanced behaviour is observed when the fusion model is employed to urban scenarios, mainly due to clearly defined sources of FAs. The proposed approach is expected to set the basis for further enhancement of data fusion mechanisms for homeland security and defence applications of c-UAV technology, taking into account at the same time the attributes of diverse environments through increased adaptability.

References

1. **A**dvanced ho**L**istic **A**dverse **D**rone **D**etection, **I**dentification and Neutralization, ALADDIN 2020 Project. `https://aladdin2020.eu/`, accessed: 2021-03-30
2. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence **41**(2), 423–443 (2018)
3. Bombini, L., Cerri, P., Medici, P., Alessandretti, G.: Radar-vision fusion for vehicle detection. In: Proceedings of International Workshop on Intelligent Transportation. vol. 65, p. 70 (2006)
4. Chang, J.R., Chen, Y.S.: Batch-normalized maxout network in network. arXiv preprint arXiv:1511.02583 (2015)
5. Christnacher, F., Hengy, S., Laurenzis, M., Matwyschuk, A., Naz, P., Schertzer, S., Schmitt, G.: Optical and acoustical uav detection. In: Electro-Optical Remote Sensing X. vol. 9988, p. 99880B. International Society for Optics and Photonics (2016)
6. Diamantidou, E., Lalas, A., Votis, K., Tzovaras, D.: Multimodal deep learning framework for enhanced accuracy of uav detection. In: International Conference on Computer Vision Systems. pp. 768–777. Springer (2019)
7. Jovanoska, S., Brötje, M., Koch, W.: Multisensor data fusion for uav detection and tracking. In: 2018 19th International Radar Symposium (IRS). pp. 1–10. IEEE (2018)
8. Kavukcuoglu, K., Sermanet, P., Boureau, Y.L., Gregor, K., Mathieu, M., Cun, Y.L.: Learning convolutional feature hierarchies for visual recognition. In: Advances in neural information processing systems. pp. 1090–1098 (2010)
9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)
10. LeCun, Y., et al.: Lenet-5, convolutional neural networks. URL: http://yann. lecun. com/exdb/lenet **20**(5), 14 (2015)
11. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11–26 (2017)

12. Manana, M., Tu, C., Owolawi, P.A.: A survey on vehicle detection based on convolution neural networks. In: 2017 3rd IEEE International Conference on Computer and Communications (ICCC). pp. 1751–1755. IEEE (2017)
13. Mészarós, J.: Aerial surveying uav based on open-source hardware and software. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **37**(1), 555 (2011)
14. Moses, A., Rutherford, M.J., Valavanis, K.P.: Radar-based detection and identification for miniature air vehicles. In: 2011 IEEE International Conference on Control Applications (CCA). pp. 933–940. IEEE (2011)
15. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In: ICML (2011)
16. Opromolla, R., Fasano, G., Accardo, D.: A vision-based approach to uav detection and tracking in cooperative applications. Sensors **18**(10), 3391 (2018)
17. Pal, S.K., Mitra, S.: Multilayer perceptron, fuzzy sets, classifiaction (1992)
18. Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z.B., Swami, A.: The limitations of deep learning in adversarial settings. In: 2016 IEEE European symposium on security and privacy (EuroS&P). pp. 372–387. IEEE (2016)
19. Samaras, S., Diamantidou, E., Ataloglou, D., Sakellariou, N., Vafeiadis, A., Magoulianitis, V., Lalas, A., Dimou, A., Zarpalas, D., Votis, K., et al.: Deep learning on multi sensor data for counter uav applications—a systematic review. Sensors **19**(22), 4837 (2019)
20. Samaras, S., Magoulianitis, V., Dimou, A., Zarpalas, D., Daras, P.: Uav classification with deep learning using surveillance radar data. In: International Conference on Computer Vision Systems. pp. 744–753. Springer (2019)
21. Shakhatreh, H., Sawalmeh, A.H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N.S., Khreishah, A., Guizani, M.: Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges. Ieee Access **7**, 48572–48634 (2019)
22. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
23. Stöcker, C., Bennett, R., Nex, F., Gerke, M., Zevenbergen, J.: Review of the current state of uav regulations. Remote sensing **9**(5), 459 (2017)
24. Thomas, A., Leboucher, V., Cotinat, A., Finet, P., Gilbert, M.: Uav localization using panoramic thermal cameras. In: International Conference on Computer Vision Systems. pp. 754–767. Springer (2019)
25. Wiatowski, T., Bölcskei, H.: A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Transactions on Information Theory **64**(3), 1845–1866 (2017)