A Preliminary Study on the Automatic Visual based Identification of UAV Pilots from Counter UAVs

Dario Cazzato[a], Claudio Cimarelli[b] and Holger Voos[c]
Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg, 29, Avenue J. F. Kennedy, 1855 Luxembourg, Luxembourg

Keywords: Pilot Identification, Skeleton Tracking, Person Segmentation, Unmanned Aerial Vehicles.

Abstract: Two typical Unmanned Aerial Vehicles (UAV) countermeasures involve the detection and tracking of the UAV position, as well as of the human pilot; they are of critical importance before taking any countermeasure, and they already obtained strong attention from national security agencies in different countries. Recent advances in computer vision and artificial intelligence are already proposing many visual detection systems from an operating UAV, but they do not focus on the problem of the detection of the pilot of another approaching unauthorized UAV. In this work, a first attempt of proposing a full autonomous pipeline to process images from a flying UAV to detect the pilot of an unauthorized UAV entering a no-fly zone is introduced. A challenging video sequence has been created flying with a UAV in an urban scenario and it has been used for this preliminary evaluation. Experiments show very encouraging results in terms of recognition, and a complete dataset to evaluate artificial intelligence-based solution will be prepared.

1 INTRODUCTION

Aerial robotics is steadily gaining attention from the computer vision research community as the challenges involved in an autonomous flying system provide a prolific field for novel applications. In particular, the video-photography coverage guaranteed by drone mobility is exploited for precision agriculture, building inspections, security and surveillance, search and rescue, and road traffic monitoring (Shakhatreh et al., 2019). Hence, the large volume of generated imagery data drives the development of more sophisticated algorithms for its analysis and of systems for real-time onboard computation (Kyrkou et al., 2018). Moreover, the popularity of Unmanned Aerial Vehicles (UAV) has exponentially increased as a business's opportunity to convert manual work into an automated process for a diverse range of industries. In confirmation of this, a recent report from PricewaterhouseCoopers (Mazur et al., 2016) identifies the businesses which will be impacted the most from the development of "drone powered solutions" in those that need high-quality (visual) data or the versatile capabilities of surveying an area. Nonetheless, in recent years, commercial drones filled the market bringing the outcome of applied computer vision research into small devices used by hobbyist as recreational tools. As a consequence of the widespread adoption of UAVs, security-related issues start to arise demanding for clear regulations and for solutions to intervene against violations of such rules. With this regard, Deloitte outlines the different risk scenarios, classified either as physical or as cyber risk, and the type of actors involved, e.g., unintentional subjects unaware of flight restrictions or deliberately malicious actors, proposing a strategy to continuously monitor the state of active countermeasures and to properly respond to a particular threatening situation (Deloitte, 2018). Hartman and Giles (Hartmann and Steup, 2013) expose the past incident caused by violations of restricted aerial space, for example, in the vicinity of airports, and warn against the critical issues implied by the large availability of UAVs for the general public. Regarding economic implications, an article of Fortune (Fortune, 2019) estimates that the recent incident occurred at Gatwick airport costs for the airlines was more than 50 million in conjunction with the cancellation of over 1.000 flights.

Hence, the search for robust and safe countermeasures against potential misuses of UAVs, e.g., flying

[a] https://orcid.org/0000-0002-5472-9150
[b] https://orcid.org/0000-0002-0414-8278
[c] https://orcid.org/0000-0002-9600-8386

Cazzato, D., Cimarelli, C. and Voos, H.
A Preliminary Study on the Automatic Visual based Identification of UAV Pilots from Counter UAVs.
DOI: 10.5220/0009355605820589
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 582-589
ISBN: 978-989-758-402-2; ISSN: 2184-4321

in protected airspace, is of critical importance. Therefore, it is not surprising that the identification of UAVs represents a very popular research topic (Korobiichuk et al., 2019), especially in defence and security domains where drones are, currently, often under real-time human control or with small levels of autonomy. While the visual detection and tracking of unauthorized UAVs are already considered in the recent state of the art research (Guvenc et al., 2018; Wagoner et al., 2017), the problem of a vision-based identification of the pilot controlling the drone still represents a missing research area.

Herein, we assume that a sensor system for the detection and tracking of UAVs intruding restricted airspace is already in place. We then propose to extend this system by a counter UAV for automatic detection and identification of the pilot(s) of the intruder UAV(s). In this work, we provide a preliminary study on how to perform the identification of the pilot with a vision system on-board of the counter UAV. The proposed approach opportunely couples two deep learning solution to extract the area and the skeleton of each person present in the scene. 2D features of joints positions are extracted and used to distinguish between the pilot and the position of other surrounding people. A first long video sequence simulating an operational scenario, with multiple people even framed at the same time, has been recorded and labelled, and a laboratory evaluation has been performed in benchmark videos with different participants and background, obtaining very encouraging results.

The rest of the manuscript is organized as follows: Section 2 reports the related work; in Section 3, the proposed approach is illustrated, and each block is detailed. Section 4 reports the scenario used for creating training and test data, used in Section 5 to validate the system, where the obtained results are discussed. Section 6 concludes the manuscript with future work.

2 STATE OF THE ART

Detection of unauthorized UAVs typically involves the usage and fusion of different signals and means, like radio-frequency (Zhang et al., 2018; May et al., 2017; Ezuma et al., 2019), WiFi fingerprint (Bisio et al., 2018), or integration of data coming from different sensors (Jovanoska et al., 2018), often requiring special and expensive hardware (Biallawons et al., 2018; May et al., 2017). Additionally, solutions based on pure image processing have been proposed exploiting recent advances in artificial intelligence and, in particular, deep learning (Carnie et al., 2006; Unlu et al., 2018; Unlu et al., 2019). If vision-based surveillance has massively been taken into account from the research community (Kim et al., 2010; Morris and Trivedi, 2008), very few works to directly identify the human operator have been proposed. Knowing who is piloting the UAV could lead to proper measures from the legal authorities, as well as having the effect of hijacking risk mitigation. These works usually consider radio frequency of the remote controller or vulnerabilities on the communication channel (Hartmann and Steup, 2013). One example is represented by GPS spoofing (Zhang and Zhu, 2017) that consists in deceiving a GPS receiver by broadcasting incorrect GPS signals; this vulnerability can be exploited to deviate the UAV trajectory (Su et al., 2016) with malicious intents, and it has been also utilized as a means to capture unauthorized flying vehicles (Kerns et al., 2014; Gaspar et al., 2018). Furthermore, modelling the pilot flight style through the sequence of radio commands has been investigated (Shoufan et al., 2018); here temporal features are exploited together with machine learning techniques to identify the pilot based on its behavioural pattern. Anyway, Shoufan et al. used behaviour as a mean for soft-biometrics without the intent of providing spacial localization. Finally, methods for visual-based action recognition (Poppe, 2010) do not consider the piloting behavior, whose presence in popular datasets, e.g., NTU-RGB+D (Liu et al., 2019) or UCF101 (Soomro et al., 2012), is absent. Therefore, from an analysis of the state of the art emerges that this specific problem has never been dealt before with a computer vision approach.

3 PROPOSED METHOD

In Figure 1, a block diagram displays an overview of the proposed pipeline. As a first step, imagery data is recorded from a camera mounted on a flying UAV. In each frame, the presence of people is detected defining the bounding box around the person body and, successively, the skeleton is estimated. Then, human joints' positions are extracted and normalized with regards to the size of the bounding box containing the body. These body features are the input for a Support Vector Machine (SVM) classifier, which has been trained on manually labelled video-sequences, to assess if the detected person represents or not a pilot. The class label is given back to the UAV in order to take proper action, like flying closer to the target, activating a tracker, etc. In the next subsections, each block will be detailed.

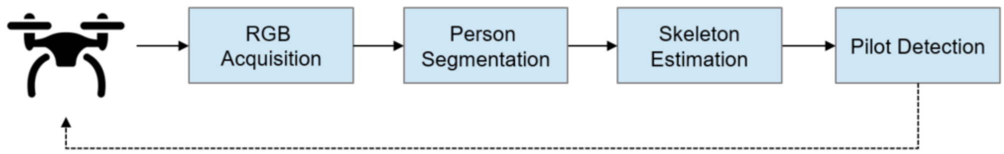


Figure 1: A block diagram of the pilot detection method. The final output is the class label that can be transmitted back to the UAV in order to take proper action, as represented by the dashed line.

3.1 RGB Acquisition

Colour images are taken from the UAV onboard camera. The scenario under consideration has one or more UAVs flying autonomously and with a pre-planned trajectory for the detection and tracking of other UAVs intruding in restricted airspace. In the case of a swarm or when multiple UAVs are operating, many data is produced for each instant of time; thus, a semi-automatic support system that automatically can identify the pilot becomes critical for taking proper countermeasures and/or defining any human intervention from the legal authorities.

3.2 Person Segmentation

People are recognized and their region is segmented by means of Mask R-CNN (He et al., 2017), an extension of the Faster R-CNN (Ren et al., 2015), i.e., a region proposal network (RPN) that shares full-image convolutional features with another network trained for detection. RPNs are fully-convolutional neural networks that take an image as input and output a set of rectangular object proposals, i.e., a group of regions of interest (RoI) with an objectness score; such regions are warped by a RoI pooling and are finally used by the detection network. Mask R-CNN introduces a binary mask for each region of interest and a RoI alignment layer that removes the harsh quantization of RoI pooling, obtaining a proper alignment of the extracted features with the input.

The output of Mask R-CNN is the list of object classes that are detected in the image and the respective mask, labelling the pixels that belong to each object instance. From the original 80 MS COCO (Lin et al., 2014) categories, we extract any occurrence of the *person* class.

Figure 2 reports some example of processed images in this first algorithmic step.

3.3 Skeleton Estimation

For each mask containing a person, a bounding box enlarged of 5 pixels in each direction is extracted and its content is processed with OpenPose (Cao et al., 2018) in order to estimate the 2D locations of anatomical keypoints for each person in the image (*skeleton*). OpenPose is a multi-stage CNN that iteratively predicts affinity fields that encode part-to-part association and confidence maps. The iterative prediction architecture refines the predictions over successive stages with intermediate supervision at each stage.

Only people that are in frontal or in a lateral view, i.e., in the range of $\pm 90^\circ$ from the frontal pose, are considered for the classification. Since the difference between a UAV pilot and a person doing other activities cannot be noticed using solely visual clues when the person is turned around, this is a realistic assumption. Hence, when the pose is in the aforementioned range, the pixel positions (2D) of the 25 joints are extracted. Following, all the correctly detected joints' positions, specified by their (x,y) pixel coordinates, are normalized inside the interval $[0-1]$ w.r.t. to the bounding box size previously obtained. Otherwise, if a joint is not detected, its (x,y) coordinates are set to -1. Finally, a feature vector of 50 elements is formed. Two outputs of this phase are shown in Figure 3.

3.4 Pilot Detection

Each feature vector is processed by an SVM classifier (Cortes and Vapnik, 1995) that outputs the label class, i.e., "*pilot*" or "*non-pilot*", for each detected person. Radial Basis Functions (RBF) is employed as the kernel. The classifier has been previously trained on six different video sequences of four pilots and two non-pilots, using manually labelled data for the class to which it corresponds (see Section 5). Thus, we used k-fold cross-validation technique (Han et al., 2011) to find the best SVM parameters. Randomized search strategy (Bergstra and Bengio, 2012) has been applied, performing k-fold cross-validation with diverse parameter's combinations by randomly sampling from the distribution of each parameter. At last, the predicted class label can be transmitted back to the UAV in order to take a proper countermeasure for the case under consideration, i.e., a sensor system for the detection and tracking of UAVs intruding restricted airspace.

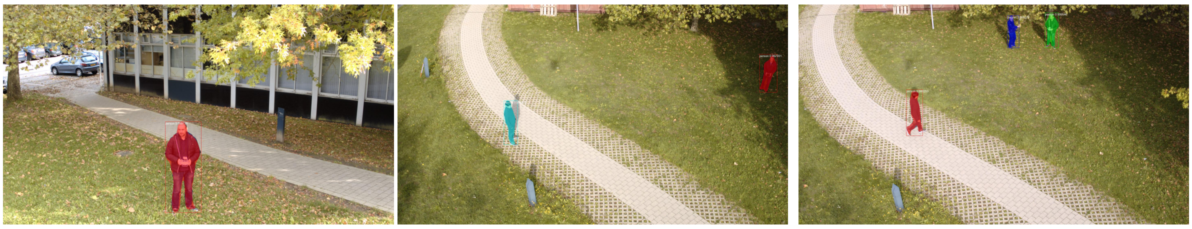


Figure 2: Outputs of the person segmentation step.

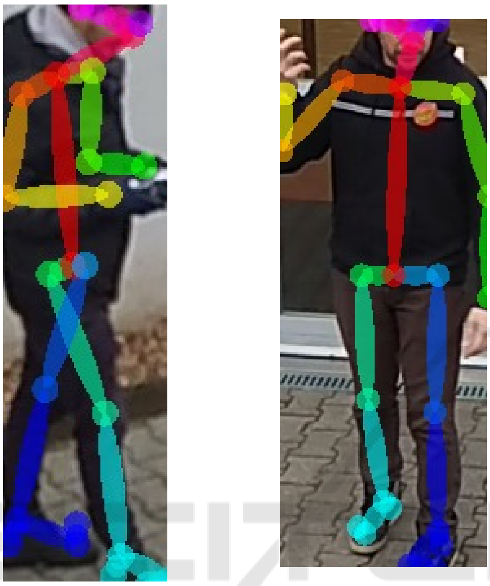


Figure 3: Two outputs of the skeleton estimation block.

4 EXPERIMENTAL SETUP

Since the objective of this work is to detect and recognize people that are possibly controlling UAVs through visual clues, the creation of a dataset is fundamental for the use of machine learning techniques. A dataset that is challenging and that pushes the classifier to generalise for a particular vision task should certainly provide different points of view and different scene situations of the subject under analysis. Because the complete pilot detection task is going to be carried out on a drone at certain levels of altitude, it was thought appropriate for the training image sequence to be filmed from a consumer camera placed at a variable height in the range of $[2.5-3.0]$ meters. Particularly, six videos with a single person present in all the frames, either a pilot or a non-pilot, were created. In Figure 4, four random frames of the aforementioned sequences are visualized to exemplify.

Therefore, to evaluate the proposed approach, a realistic test sequence in an urban scenario has been prepared. In particular, a DJI M600 mounting a stabilized DJI Zenmuse X5R camera has been used and manually piloted, changing its orientation and height, recording different parts of the area while exploring it. The UAV is flying in the range $[2.5-10.0]$ meters. The background varies and it includes buildings, parked cars, a field and an alleyway. Different people without constraints about the behaviour, appearance and clothes are present in the scene. Different people can be present at a given frame; among them, two pilots holding a controller. The video has been created using Mask-R CNN to extract the bounding box around detected people and manually assigning a binary label representing the action of (non-)piloting.

The size of the training and test dataset is summarized in Table 1. Note that, while the classes are almost completely balanced among the training sequences, in the case of the test sequence this number is different from the number of images containing the video since multiple, as well as or no person, can be present in the scene at a given instant.

Table 1: The number of images used for each class label in the training and test sequences.

Class label	Training	Test
Pilot	3699	1879
Non-pilot	2960	663
Total	6659	2542

About implementation details, all the code has been written in Python and with the support of OpenCV utility functions. The Mask R-CNN implementation in (Abdulla, 2017) has been used for person detection and segmentation. The Python wrapper of OpenPose (Cao et al., 2018) with pre-trained weights for 25-keypoint body/foot estimation has been employed for the body-joint position extraction. The final phase of the classification adopts the implementation of SVM found into Scikit-Learn (Pedregosa et al., 2011). The software has been executed on an Intel Xeon(R) CPU E3-1505M v6 @3.00GHz x8 with an NVIDIA Quadro M1200 during all the phases.

Figure 4: Four images that belong to the training set. Note that it is asked to the same actor to behave both as pilot and non-pilot in different video sequences.

5 EXPERIMENTAL RESULTS

First of all, an evaluation of the person segmentation and the skeleton estimation modules is provided. In the training sequences, each person appearance has correctly detected by Mask-R CNN (100% detection rate). For each RoI containing a person, a skeleton has always been extracted by OpenPose (again, 100% detection rate). It was not possible to establish a ground truth of the skeleton, thus the obtained information has been directly used for training. In the sequence for tests, some misdetection occurs, mainly due to partial and/or too top-angled views.

Results of this experimental evaluation are summed up by Table 2. The usage of Mask R-CNN upstream from OpenPose had the advantage of drastically reduced false positives, and for two reasons: from one side, it reduced false positives of Mask-R CNN, since no skeleton data was found on such region. From the other side, we tested the usage of OpenPose without any previous step, and it generated a significant quantity of false positives. At this purpose, it can be observed that all false positives generated by Mask-R CNN have been removed by the skeleton estimation phase. Samples of these patches are reported in Figure 5. On the other side, with the proposed pipeline, it is not possible to restore false negatives. Anyway, they occurred only in 30 frames, related to the visual appearance of the three patches reported in Figure 6; in particular, these cases are always related to positions close to the border of the image, and there are 22 missed frames for Person #1 (Figure 6a), 8 frames for Person #2 (Figure 6b), 2 frames for Person #3 (Figure 6c). Finally, there are few cases of multiple skeleton estimation instead of one. In this case, simple ratio and dimension filters remove the majority of them, while the skeleton related to the framed person results noisy. Notwithstanding the pre-filtering of the dataset, the removal of all noisy skeletons from the dataset is not ensured. The solution can work also in the case of multiple people with overlapping and occlusions, as shown in Figure 7.

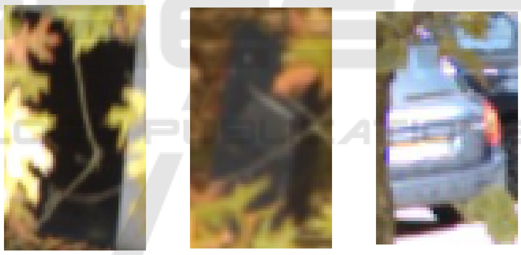
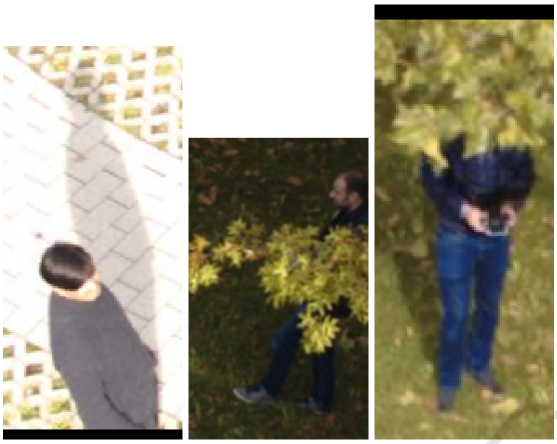


Figure 5: Examples of false positives detected by Mask R-CNN. The skeleton estimation step removed all of them.

The second experimental phase aims at evaluating the classification performance. K-fold cross-validation has been performed with k=5 to validate a

Table 2: Analysis of errors of first algorithmic steps for the test sequence. Before the classification, there are 0 false positives and 30 false negatives out of 2542 bounding boxes (1.18% of errors).

	After person segmentation	After skeleton estimation
False positives	53	0
False negatives	30	30



(a) Person #1 (b) Person #2 (c) Person #3

Figure 6: Three frames in which no person was detected.

Figure 7: An example of overlapping bounding box; also, in this case, the two different skeletons were estimated.

choice of C and γ for tuning the SVM. Hence, we sample 500 values from two different distributions to randomly search the parameters space for the best combinations. In particular, C was extracted from a uniform distribution inside the interval $[0,3]$ and γ from an exponential distribution with $\lambda = 0.02$. Results of the random search gave the following set of parameters, which obtained about 90% of accuracy on average between the 5 folds:

- $C = 2.6049$
- $\gamma = 0.02759$

Results obtained on the test data introduced in Section 4 are illustrated in Table 3 and the Receiver Operating Characteristic (ROC) curve is reported in Figure 8.

In particular, a correct detection occurs around 82% of the time, even using different video sequences from training and test, recorded at different heights, sensors and background. An example of correct classification of pilots and non-pilots is reported in Figure 9a. Also the case of the two pilots whose appearance presents an overlapping (see Figure 7) has been correctly classified.

Table 3: Classification performance for the pilot detection task.

	Precision	Recall	F1-Score
Non-pilot	0.61	0.90	0.72
Pilot	0.96	0.79	0.87
Accuracy		0.82	

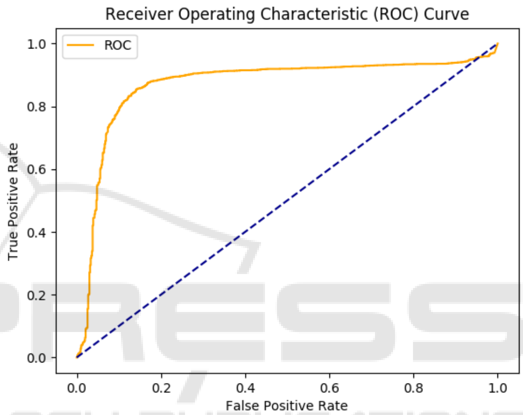
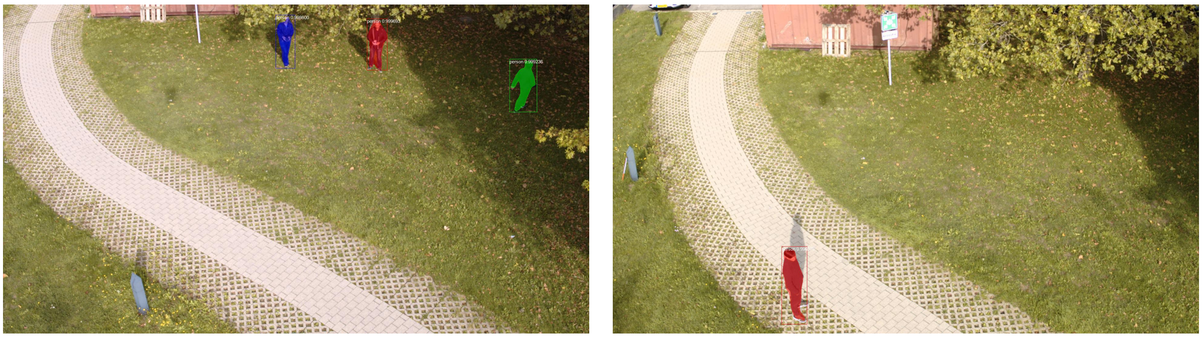


Figure 8: ROC curve results of the probabilistic SVM with Platt's method (Platt et al., 1999).

Figure 9b reports an example of misclassification. In general, the majority of errors is due to a noisy detection of the skeleton and/or with missing data of OpenPose when the UAV is flying close to the maximum height, as well as partial and occluded views. Surprisingly, possible wrong perspective due to operating with 2D data is not representing an issue, probably as consequence of the operational interval for the orientation range. Indeed, frames in which a person is holding a tablet or a mobile could be misinterpreted, but this raises a new problem since a controller for a UAV could be also represented by a mobile or a tablet. In this preliminary work, we decided to do not consider this case; adding a new class, i.e., *poten-*



(a) (b)

Figure 9: Examples of the classification results: a) Two pilots (blue and red mask) and one non-pilot (green mask) correctly classified; b) A non-pilot wrongly classified as pilot.

tial pilot, depending on the specific hold object will be evaluated. About the neural networks employed by the system for detecting a person and its pose, we have chosen Mask-R CNN for person segmentation even if it is not the state of the art solution in terms of real-time performance; this is motivated since our goal of this preliminary study was to evaluate the feasibility of the problem under consideration, thus we have chosen a state of the art network in terms of detection performance.

6 CONCLUSION

In this work, a fully autonomous pipeline to process images from a flying UAV to recognize the pilot in challenging and realistic environment has been introduced. The system as been evaluated with a dataset of images taken from a flying UAV in urban scenario, and the feasibility of an approach for piloting behavior classification based on human body joint extracted features has been evaluated. The system obtained a classification rate of 82% after being trained with different video sequences, with large precision in pilot detection and a few generations of false positives.

In the future work these results will be exploited in order to propose a faster and complete pipeline that can be directly processed on-board and with limited hardware capabilities, also introducing spatio-temporal information and data association schemes. The integration of a depth sensor to increase performance will also be investigated. Finally, we will propose a complete and labelled dataset composed by different scenes of UAVs flying at different altitudes and operative scenarios, and it will be made publicly available.

REFERENCES

Abdulla, W. (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.

Biallawons, O., Klare, J., and Fuhrmann, L. (2018). Improved uav detection with the mimo radar mira-cle ka using range-velocity processing and tdma correction algorithms. In *2018 19th International Radar Symposium (IRS)*, pages 1–10. IEEE.

Bisio, I., Garibotto, C., Lavagetto, F., Sciarrone, A., and Zappatore, S. (2018). Unauthorized amateur uav detection based on wifi statistical fingerprint analysis. *IEEE Communications Magazine*, 56(4):106–111.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.

Carnie, R., Walker, R., and Corke, P. (2006). Image processing algorithms for uav" sense and avoid". In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 2848–2853. IEEE.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Deloitte (2018). Unmanned aircraft systems (uas) risk management: Thriving amid emerging threats and opportunities.

Ezuma, M., Erden, F., Anjinappa, C. K., Ozdemir, O., and Guvenc, I. (2019). Micro-uav detection and classification from rf fingerprints using machine learning techniques. In *2019 IEEE Aerospace Conference*, pages 1–13. IEEE.

Fortune (2019). Gatwicks december drone closure cost airlines \$ 64.5 million.

Gaspar, J., Ferreira, R., Sebastião, P., and Souto, N. (2018). Capture of uavs through gps spoofing. In *2018 Global Wireless Summit (GWS)*, pages 21–26. IEEE.

Guvenc, I., Koohifar, F., Singh, S., Sichitiu, M. L., and Matolak, D. (2018). Detection, tracking, and interdiction for amateur drones. *IEEE Communications Magazine*, 56(4):75–81.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hartmann, K. and Steup, C. (2013). The vulnerability of uavs to cyber attacks-an approach to the risk assessment. In *2013 5th international conference on cyber conflict (CYCON 2013)*, pages 1–23. IEEE.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

Jovanoska, S., Brötje, M., and Koch, W. (2018). Multisensor data fusion for uav detection and tracking. In *2018 19th International Radar Symposium (IRS)*, pages 1–10. IEEE.

Kerns, A. J., Shepard, D. P., Bhatti, J. A., and Humphreys, T. E. (2014). Unmanned aircraft capture and control via gps spoofing. *Journal of Field Robotics*, 31(4):617–636.

Kim, I. S., Choi, H. S., Yi, K. M., Choi, J. Y., and Kong, S. G. (2010). Intelligent visual surveillancea survey. *International Journal of Control, Automation and Systems*, 8(5):926–939.

Korobiichuk, I., Danik, Y., Samchyshyn, O., Dupelich, S., and Kachniarz, M. (2019). The estimation algorithm of operative capabilities of complex countermeasures to resist uavs. *Simulation*, 95(6):569–573.

Kyrkou, C., Plastiras, G., Theocharides, T., Venieris, S. I., and Bouganis, C.-S. (2018). Dronet: Efficient convolutional neural network detector for real-time uav applications. In *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 967–972. IEEE.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

Liu, J., Shahroudy, A., Perez, M. L., Wang, G., Duan, L.-Y., and Chichung, A. K. (2019). Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. *IEEE transactions on pattern analysis and machine intelligence*.

May, R., Steinheim, Y., Kvaløy, P., Vang, R., and Hanssen, F. (2017). Performance test and verification of an off-the-shelf automated avian radar tracking system. *Ecology and evolution*, 7(15):5930–5938.

Mazur, M., Wisniewski, A., and McMillan, J. (2016). Clarity from above: Pwc global report on the commercial applications of drone technology. *Warsaw: Drone Powered Solutions, PriceWater house Coopers*.

Morris, B. T. and Trivedi, M. M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. *IEEE transactions on circuits and systems for video technology*, 18(8):1114–1127.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Platt, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.

Shakhatreh, H., Sawalmeh, A. H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N. S., Khreishah, A., and Guizani, M. (2019). Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges. *IEEE Access*, 7:48572–48634.

Shoufan, A., Al-Angari, H. M., Sheikh, M. F. A., and Damiani, E. (2018). Drone pilot identification by classifying radio-control signals. *IEEE Transactions on Information Forensics and Security*, 13(10):2439–2447.

Soomro, K., Zamir, A. R., and Shah, M. (2012). Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

Su, J., He, J., Cheng, P., and Chen, J. (2016). A stealthy gps spoofing strategy for manipulating the trajectory of an unmanned aerial vehicle. *IFAC-PapersOnLine*, 49(22):291–296.

Unlu, E., Zenou, E., and Riviere, N. (2018). Using shape descriptors for uav detection. *Electronic Imaging*, 2018(9):1–5.

Unlu, E., Zenou, E., Riviere, N., and Dupouy, P.-E. (2019). Deep learning-based strategies for the detection and tracking of drones using several cameras. *IPSJ Transactions on Computer Vision and Applications*, 11(1):7.

Wagoner, A. R., Schrader, D. K., and Matson, E. T. (2017). Survey on detection and tracking of uavs using computer vision. In *2017 First IEEE International Conference on Robotic Computing (IRC)*, pages 320–325. IEEE.

Zhang, H., Cao, C., Xu, L., and Gulliver, T. A. (2018). A uav detection algorithm based on an artificial neural network. *IEEE Access*, 6:24720–24728.

Zhang, T. and Zhu, Q. (2017). Strategic defense against deceptive civilian gps spoofing of unmanned aerial vehicles. In *International Conference on Decision and Game Theory for Security*, pages 213–233. Springer.