Proceedings of The 2023 IEEE International
Conference on Real-time Computing and Robotics
July 17-20, 2023, Datong, China

Real-time UAV Localization and Tracking in Multi-Weather Conditions using Multispectral Image Analysis*

Yuxiang Lin, Xiaojiang Peng, Jiahe Yu, Wei Chen, Yan Wu, and Hu Liu

***Abstract*— With the increasing availability of unmanned aerial vehicles (UAV), their potential misuse has become a serious concern, posing a threat to public security. Existing tracking methods have limitations in detecting UAV effectively due to their small size, high speed, complex flight patterns, and complicated flying background. To address this problem, we propose a UAV localization and tracking method that uses multispectral images captured by specific hardware, which enhances the detection process by allowing for greater visibility in challenging weather conditions. The proposed method combines the YOLOv5 detection algorithm with the KCF tracking algorithm to provide a reliable solution for preventing potential misuse of UAV and enhancing public safety.**

Experimental results demonstrate that the proposed method provides a reliable solution for UAV localization and tracking in various weather conditions. The method was found to improve the inference speed compared to a single YOLOv5 model, demonstrating its potential for real-time UAV tracking and control. By combining multispectral image analysis, detection algorithms, and tracking algorithms, the proposed method provides an effective solution for preventing the potential misuse of UAV and enhancing public safety. This research presents a promising direction for future studies on UAV tracking and control.

***Index Terms*— UAV, multispectral, real-time, YOLOv5, KCF**

I. INTRODUCTION

Unmanned Aerial Vehicles (UAV), also known as drones, have become increasingly popular due to their low cost, ease of operation, and ability to capture high-quality images [1][2]. The use of UAV has been particularly significant in the field of aerial photography, where high-quality images and videos can be captured from above. In agriculture, UAV have been used for crop monitoring and analysis, allowing farmers to better understand the health of their crops and make informed decisions. In surveillance, UAV can be used for border control, disaster management, and law enforcement operations. Additionally, UAV have proven useful in search and rescue missions by providing aerial views of the affected areas and helping to locate survivors [3][4][5].

However, there are concerns regarding the misuse of UAVs, particularly in sensitive areas. UAV have been known to interfere with the operations of airports and other critical infrastructure, putting lives and property at risk. Privacy concerns have also been raised, as UAV can be used for intrusive surveillance, invading individuals' personal space [6][7][8]. To address these concerns, UAV localization and tracking technology has become critical solution. It involves monitoring the location and movement of UAV, and is essential in ensuring the safety and security of individuals and property.

Various methods have been proposed for tracking Unmanned Aerial Vehicles. These methods include GPS-based tracking, radar systems, acoustic tracking, and visual tracking. GPS-based tracking relies on satellite signals to determine the UAV's location [9]. However, this method may not be effective if the UAV does not use GPS or uses a different frequency. Additionally, accessing or manipulating GPS data of other users may be illegal or unethical.

Radar systems, on the other hand, can detect UAV by measuring their position, velocity, acceleration, and direction [10][11]. However, they may be affected by interference from other sources of radio waves and may be expensive and complex to install and maintain. Acoustic tracking is another method that uses microphones to detect the sound of the UAV. However, this method has a limited range and may not be able to distinguish between different UAV of the same model or between UAV and other sources of noise [12][13].

Visual tracking methods, such as using cameras and image processing techniques, have proven to be highly accurate and cost-effective. This method can provide high-resolution images of the UAV and can be used to track the UAV's movement. Compared to the aforementioned methods, visual tracking has the outstanding advantage of highly accurate detection, especially for UAV details. However, it may not perform well in low-visibility conditions, such as fog or darkness.

To address this limitation, we propose a method that utilizes multispectral images to detect and track UAVs in multi-weather conditions. Specifically, we use RGB images to detect UAVs in normal weather conditions and thermal infrared images to enhance detection in low-visibility conditions such as darkness, clutter, or occlusion. By combining

*This work was supported by Water Conservancy Science and Technology Innovation Project of Guangdong Province(no. 2022-03)

Yuxiang Lin is with Shenzhen Technology University and Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China (e-mail: yx.lin2@siat.ac.cn).

Xiaojiang Peng is with Shenzhen Technology University, Shenzhen, 518000, China (e-mail: pengxiaojiang@sztu.edu.cn).

Jiahe Yu is with Shenyang Aerospace University and Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China (e-mail: jh.yu@siat.ac.cn).

Yan Wu is with Xi'an Institute of Applied Optics, Xi'an, 710065, China (e-mail: 1195442231@qq.com).

Wei Chen is with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China (corresponding author to provide phone: 0755-86392483; fax: 0755-86392299; e-mail: chenwei@siat.ac.cn).

Hu Liu is with Xi'an Institute of Applied Optics, Xi'an, 710065, China (corresponding author to provide phone: 029-88288135; e-mail: bluesinwind@126.com).

these two spectra of images, we can improve the accuracy and reliability of UAV detection and tracking, making it suitable for a wide range of applications.

In this paper, we present a novel approach for robust UAV localization and tracking using Multispectral Image Analysis (RGB and Thermal infrared images). We have created a dataset of multispectral UAV images captured at a resolution of 1920x1080. We combine the You Only Look Once [14] version 5 (YOLOv5) object detection algorithm with the Kernel Correlation Filter Tracking Algorithm (KCF) [15] to develop an approach that is capable of real-time tracking with high accuracy, cost-effectiveness, and the ability to retrack lost targets.

Our experimental results demonstrate that our proposed algorithm is a promising solution to the issue of anti-UAV, with the potential to significantly improve public safety and security. In summary, our contributions are as follows:

- We propose a novel method by using multispectral image analysis to solve the limitation of visual tracking methods on low-visibility at a single RGB image.
- We have created a dataset consisting of multispectral UAV images captured at a resolution of 1920x1080.
- We have devised an approach for real-time tracking that integrates the strengths of both the YOLOv5 object detection algorithm and the KCF tracking algorithm while compensating for their respective weaknesses. Our method is characterized by high precision, affordability, and the capability to recover lost targets.

II. Related Work

A. YOLOv5 Object Detection Algorithm

Object detection is a crucial task in various fields, including unmanned aerial vehicle (UAV) pest control. The YOLOv5 algorithm is a popular one-stage object detection algorithm that offers advantages over previous versions, such as smaller mean weights, shorter training time, and faster detection speed. YOLOv5 uses a detection strategy that divides the input image into multiple grids, with each grid responsible for predicting the object's position if it contains the target object. The final output is the predicted box with the highest intersection over union (IoU) with the ground-truth box.

The YOLOv5 model comprises four parts: the input end, Backbone, Neck, and Head. The input end includes Mosaic data augmentation, which randomly scales, crops, and arranges four images. The Backbone is the feature extraction part, which includes convolutional layers, C3, and SPPF structures. The convolutional layers encapsulate three functions: grouped convolution, batch normalization (BN), and SiLU activation function. The C3 module simplifies the previous BottleneckCSP structure, enhancing the model's ability to capture features. The SPPF structure replaces the spatial pyramid pooling (SPP) structure, which increases the forward and backward computation speeds by about 1.5 times.

The Neck uses the feature pyramid network (FPN) and path aggregation network (PAN) structure to combine the conventional FPN layer with the bottom-up feature pyramid, which fuses the extracted semantic features and position features. Furthermore, it performs feature fusion between the backbone layer and the detection layer, allowing the model to obtain more diverse feature information. The Head outputs the prediction results.

However, the computational speed of YOLOv5 heavily depends on the computing units used. For instance, when using a GPU for inference, the algorithm can achieve a processing speed of up to 65 frames per second, whereas using only a CPU results in an inference speed of approximately 12 frames per second. In UAV pest control tasks, real-time detection and tracking, as well as detection accuracy, are equally important. Thus, it is essential to explore and develop new methods that can achieve real-time detection and tracking with high accuracy while taking into account the limited computing resources available.

B. Kernel Correlation Filter Tracking Algorithm

The KCF algorithm trains a filter using ridge regression, assuming that the training samples are x_i. Its objective is to find a target classifier $f(x_i) = w^T x_i$, which minimizes the mean squared error function between the expected output y_i of the filter and the training samples x_i. Specifically, the algorithm seeks to minimize the expression as follow:

$$\min_w \sum_i (f(x_i) - y_i) + \lambda ||w||^2. \tag{1}$$

The symbol λ represents the regularization parameter. The optimal solution for the weight coefficients of the classifier w is given by:

$$w = (X^T X + \lambda I)^{-1} X^T y. \tag{2}$$

In the formula: X is a data matrix, I is an identity matrix, and y represents the expected target regression matrix.

Using the properties of cyclic matrix and Fourier transform, the solution of Equation (1) in the frequency domain is:

$$\widehat{w} = \frac{\hat{x} \odot \hat{y}}{\hat{x} \odot \hat{x}^* + \lambda}. \tag{3}$$

In the formula: $\odot$ is the dot product of elements; $\hat{x}$ is the discrete Fourier transform of x; $\hat{x}^*$ is the complex conjugate of $\hat{x}$.

In order to solve the nonlinear problem, the kernel function is introduced into the ridge regression method: $k(x, z') = \langle \varphi(x), \varphi(z') \rangle$, Then the weight coefficient w is expressed as $w = \sum_i \alpha_i \varphi(x_i)$, Where $\varphi(x)$ refers to mapping the sample x_i to a high-dimensional space.

Finding the optimal value of w is equivalent to finding the optimal value of α, where α is calculated as follows:

$$\alpha = (K + \lambda I)^{-1} y. \tag{4}$$

In the formula: Where: K is a kernel matrix composed of elements $K_{i,j} = k(x_i, x_j)$.

Based on the loop structure and fast Fourier transform (FFT), the matrix α equals:

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda}. \tag{5}$$

In the formula: k^{xx} is a vector composed of elements in the first row of kernel matrix k.

The output response is calculated as:

$$f(z) = F^{-1}(\hat{k}xz \odot \hat{\alpha}). \tag{6}$$

In the formula: F^{-1} is the inverse discrete Fourier transform; $f(z)$ is the output response value, and its maximum value is the position of the target to be tracked.

The KCF target tracking algorithm uses linear interpolation and updates the model with a fixed learning rate. The update method is as follows:

$$\begin{cases} \alpha_t = (1-\gamma)\alpha_{t-1} + \gamma\alpha_t' \\ x_t = (1-\gamma)x_{t-1} + \gamma x_t'. \end{cases} \tag{7}$$

In the formula: γ is the learning rate; α_t, x_t and α_{t-1}, x_{t-1} represent the parameters and target templates of the t frame and the $t-1$ frame respectively.

The updated formula of the KCF target tracking algorithm, as shown in Equation (7), utilizes a fixed learning rate γ. However, when the target encounters occlusion, the model update based on the fixed learning rate may lead to the occluding feature being learned, resulting in the occluding object being incorrectly identified as the tracking target in subsequent frames, ultimately leading to tracking failure. Therefore, the KCF algorithm suffers from a significant weakness in its assumption that the tracking of a target will be successful after its initialization. This presumption results in the updating of the detection model based solely on the previous frame, which can result in drift in the template due to improper sampling and model updates in scenarios where the target is moving swiftly, hidden, or undergoing significant transformations. The issue is further exemplified in Figure 1.

III. PROPOSED APPROACH

Unmanned aerial vehicles (UAV) are increasingly being misused for unauthorized surveillance, which poses a significant threat to privacy and security. Therefore, there is a growing need for methods to detect, locate and track such UAV. However, selecting an appropriate algorithm for object tracking is challenging due to the trade-off between accuracy and computational speed. To address this challenge, we propose a hybrid approach that combines the strengths of the YOLOv5 and KCF algorithms.

Our proposed approach for UAV localization and tracking involves the following steps:

1) **Read the video frames to be detected:** The first step is to read the multispectral video frames captured by the sensor. Our system uses both thermal infrared images and RGB images to detect objects in low-visibility conditions and normal conditions, respectively.
2) **Extract features from the video frames using the YOLOv5 algorithm and determine the location of the detection box:** We employ the YOLOv5 algorithm to extract features from each frame of the video stream. This involves running the YOLOv5 algorithm on each frame to identify the location of objects in the frame. We then use the information from the YOLOv5 algorithm to determine the location of the detection box for each object.
3) **Pass the detection box to the KCF algorithm and initialize the tracker:** Once we have determined the location of the detection box for each UAV in the video frame, we pass this information to the KCF algorithm. We use the detection box as the initial location of the object to be tracked and initialize the KCF tracker.
4) **Use the KCF algorithm to track the object in real-time:** With the KCF tracker initialized, we use the algorithm to track the object in real-time as it moves through subsequent frames of the video stream. The KCF algorithm is known for its fast processing speeds, which allows us to track objects in real-time.
5) **Switch the mode to thermal infrared images if low-visibility conditions are detected:** In the event that low-visibility conditions are detected, we switch the mode to thermal infrared images to improve object detection accuracy and repeat step 2 to re-detect the UAV. We continue tracking objects using the KCF algorithm and thermal infrared images until normal visibility is restored.
6) **If the KCF algorithm loses track of the object, repeat step 2:** If the KCF algorithm loses track of the object being tracked, we repeat step 2 to re-detect the UAV and determine its new location. We then pass this information back to the KCF tracker to continue the tracking process.

Overall, our proposed approach combines the high accuracy of the YOLOv5 algorithm with the fast processing speeds of the KCF algorithm to enable real-time object tracking in multispectral video streams. Our approach also includes the ability to switch between thermal infrared and RGB images, depending on the visibility conditions, to ensure accurate object detection.

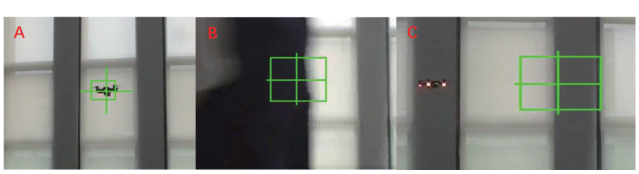

Fig. 1. The picture shows a sequence of events captured from left to right. In the first frame, the KCF algorithm is used to lock onto a UAV. However, in the second frame, a person passes in front of the camera, causing the UAV to become obscured and lose track.

IV. EXPERIMENTS

A. YOLOv5 Model Training

We have compiled a dataset comprising RGB/thermal infrared images to facilitate the localization and tracking of UAV that may be exploited for surveillance purposes. The dataset includes a total of 3000 images, which are comprised of both RGB and thermal infrared imagery. Furthermore, we have utilized the LabelMe tool to manually annotate the dataset.

The PyTorch deep learning framework was employed to train the YOLOv5 model on the annotated dataset. The model

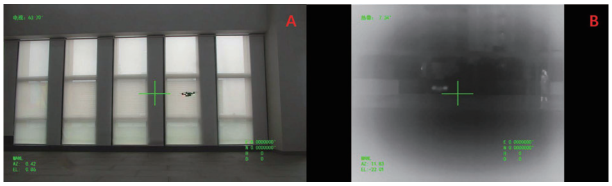

Fig. 2. This is a sample of the dataset of RGB (Right) and thermal infrared images (Left) collected for the purpose of localizing and tracking misused UAV.

was trained for 100 epochs with a batch size of 128, a learning rate of 0.01, and the stochastic gradient descent (SGD) optimizer, using a single NVIDIA V100 graphics card. To prevent overfitting, early stopping with a patience of 10 was also implemented.

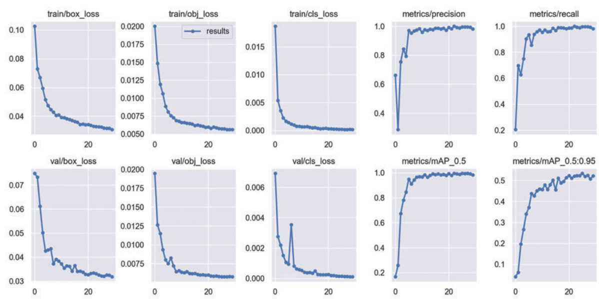

Fig. 3. Training results of the YOLOv5 model.

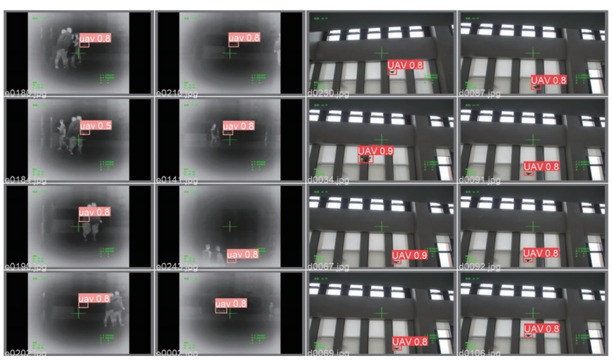

Fig. 4. Detection results on the validation set using the YOLOv5 model.

Figure 3 shows the training results of the YOLOv5 model, with the loss decreasing significantly over the training epochs. The training loss reached a plateau after around 40 epochs, indicating that the model had converged. We also evaluated the performance of the YOLOv5 model on a separate validation set, consisting of 300 images that were not used in the training process. The detection results on the validation set are shown in Figure 4, which demonstrate that the YOLOv5 model was able to accurately detect the objects in the images.

Overall, our YOLOv5 model achieved a mean average precision (mAP) of 0.9765 on the validation set, indicating that it is suitable for detecting objects in video streams.

TABLE I

THE YOLOV5 PERFORMANCE AT THE MULTISPECTRAL DATASET.

Spectral	CLS Loss	OBJ Loss	BOX Loss	mAP
RGB	0.0004	0.0195	0.02587	0.9873
Infrared	0.0005	0.0197	0.02500	0.9657

B. KCF Algorithm Implementation

The Kernelized Correlation Filter (KCF) algorithm is a popular object-tracking algorithm due to its efficiency and effectiveness. In our video tracking system, we use the KCF algorithm to track the objects of interest after they have been detected by the YOLOv5 algorithm.

To implement the KCF algorithm, we use the OpenCV library in Python. Specifically, we initialize the KCF tracker with the bounding boxes provided by the YOLOv5 algorithm, which serve as the initial positions of the tracked objects. Once the tracker has been initialized, it updates the position of the object in each subsequent frame in real-time.

The KCF algorithm is prone to losing track of the object when it moves too quickly or when it is occluded by other objects. In such cases, the YOLOv5 algorithm is used to re-detect the object and provide a new bounding box, which is then used to re-initialize the KCF tracker. This process is repeated until the object is successfully tracked.

However, the degree of confidence of KCF algorithm in the OpenCV-python library is not provided. To address this limitation, we propose a methodology that involves monitoring the motion of the bounding box returned by the KCF algorithm. Specifically, if the target is lost, the bounding box will appear to be stationary. To address this, we introduce a patient frame parameter that is set to 5 frames. If the bounding box remains stationary for up to 5 frames, we assume that the KCF algorithm has lost the target and initiate a re-detection using YOLOv5. Our experimental results demonstrate that the combination of YOLOv5 and KCF algorithms yields robust object tracking in real-time video streams.

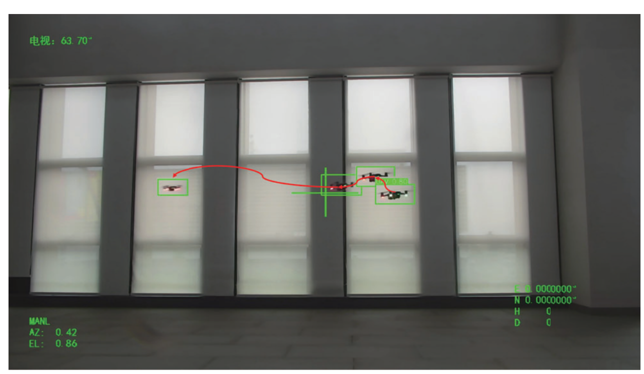

Fig. 5. Example of using proposed method to localize and track UAV. The first bounding box with 'UAV' label is detected by YOLOv5 algorithm, and the following bounding boxes is tracked by KCF tracking algorithm as proposed above.

TABLE II

EXPERIMENT RESULT OF FRAME PER SECOND WHEN DETECTING.

Method	Frame Per Second (FPS)
YOLOv5	12.5FPS
Proposed	31FPS

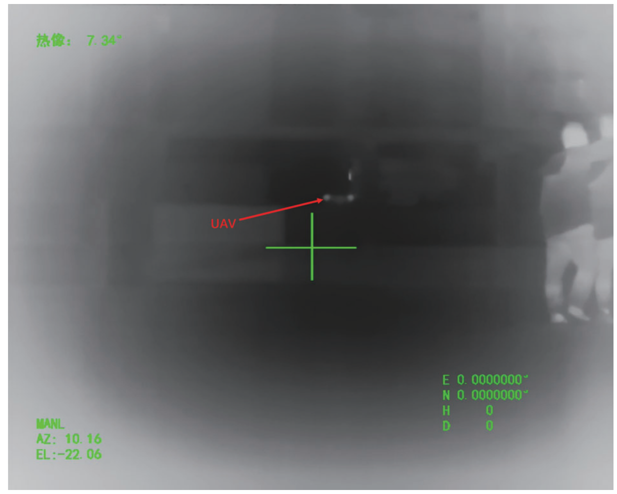

Fig. 6. UAV in an infrared image, with little detail.

V. CONCLUSIONS

In conclusion, our study has made substantial contributions to the field of unmanned aerial vehicle (UAV) localization and tracking. Firstly, we have produced a high-quality thermal infrared UAV dataset that can be employed for future research in this domain. Secondly, we have presented a novel approach that combines both RGB and infrared images to detect UAV in diverse settings, including low-light, cluttered, or obscured conditions. This method has demonstrated its robustness and efficacy, surpassing existing techniques in tracking accuracy and real-time performance. Lastly, through the integration of the YOLOv5 object detection algorithm and the KCF tracking algorithm, we have devised a method that can track UAV with high accuracy and retrack lost targets in real-time. These contributions have the potential to enhance public safety and security by preventing UAV misuse.

However, we must acknowledge that there exist other tracking methods that merit investigation. For example, deep learning-based methods like DeepSORT [16], ByteDance [17], QDTrack [18] have shown promising results in object tracking tasks and could be adapted for UAV tracking. Additionally, other traditional tracking methods such as Mean Shift [19], Particle Filter [20] and Kalman Filter [21] may also be beneficial in different scenarios.

We highly encourage researchers to explore target detection algorithms that are specifically tailored for infrared images. This is because the loss of detail in infrared images is much greater than that in RGB images, which means that algorithms that perform well on RGB images may not perform well on infrared images. Therefore, it is important to develop algorithms that are optimized for detecting targets in infrared images.

Thus, we advocate for researchers to explore and compare different tracking methods in order to further enhance the accuracy and efficiency of UAV tracking and control. By continuously pushing the limits of research in this field, we can develop more effective and robust solutions that will strengthen public safety and security.

REFERENCES

[1] V. Kharchenko and I. Chyrka, "Detection of airplanes on the ground using yolo neural network," in *2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET)*, 2018, pp. 294–297.

[2] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on uav communications for 5g and beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327–2375, 2019.

[3] M. Abdelkader, S. Güler, H. Jaleel, and et al., "Aerial swarms: Recent applications and challenges," *Current Robotics Reports*, vol. 2, pp. 309–320, 2021.

[4] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *IEEE robotics & automation magazine*, vol. 19, no. 3, pp. 46–56, 2012.

[5] P. Doherty and P. Rudol, "A uav search and rescue scenario with human body detection and geolocalization," in *AI 2007: Advances in Artificial Intelligence: 20th Australian Joint Conference on Artificial Intelligence, Gold Coast, Australia, December 2-6, 2007. Proceedings 20*. Springer, 2007, pp. 1–13.

[6] Anonymous, "A review on object detection in unmanned aerial vehicle surveillance," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 215–228, 2021. [Online]. Available: https://doi.org/10.1016/j.ijcce.2021.11.005

[7] E. Semsch, M. Jakob, D. Pavlicek, and M. Pechoucek, "Autonomous uav surveillance in complex urban environments," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 2. IEEE, 2009, pp. 82–85.

[8] Z. Li, Y. Liu, R. Hayward, J. Zhang, and J. Cai, "Knowledge-based power line detection for uav surveillance and inspection systems," in *2008 23rd International Conference Image and Vision Computing New Zealand*. IEEE, 2008, pp. 1–6.

[9] P. M. Wyder, Y.-S. Chen, A. J. Lasrado, R. J. Pelles, R. Kwiatkowski, E. O. A. Comas, R. Kennedy, A. Mangla, Z. Huang, X. Hu, Z. Xiong, T. Aharoni, T.-C. Chuang, and H. Lipson, "Autonomous drone hunter operating by deep learning and all-onboard computations in gps-denied environments," *PLOS ONE*, vol. 14, no. 11, pp. 1–18, 11 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0225092

[10] M. C. Eyler, B. Anderson, C. K. Peterson, K. Warnick, and T. Mclain, "Decentralized uav tracking with networked radar systems," in *AIAA Scitech 2021 Forum*, 2021.

[11] E. D. S. Musa, r. s. a. raja abdullah, A. Sali, A. Ismail, and A. Salah, "A review of copter drone detection using radar system," vol. 12, pp. 16–38, 03 2019.

[12] A. Bernardini, F. Mangiatordi, E. Pallotti, and L. Capodiferro, "Drone detection by acoustic signature identification," *Electronic Imaging*, vol. 2017, pp. 60–64, 01 2017.

[13] "Australia's continental-scale acoustic tracking database and its automated quality control process | Scientific Data." [Online]. Available: https://www.nature.com/articles/sdata2017206

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[15] Y. Li, J. Zhu, *et al.*, "A scale adaptive kernel correlation filter tracker with feature integration." in *ECCV workshops (2)*, vol. 8926. Citeseer, 2014, pp. 254–265.

[16] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3645–3649.

[17] Y. Zhang, P. Sun, Y. Jiang, D. Yu, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," 2021.
[18] J. Pang, L. Qiu, X. Li, H. Chen, Q. Li, T. Darrell, and F. Yu, "Quasi-dense similarity learning for multiple object tracking," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 164–173.
[19] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
[20] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE signal processing magazine*, vol. 20, no. 5, pp. 19–38, 2003.
[21] G. Welch, G. Bishop, *et al.*, "An introduction to the kalman filter," 1995.