Real-time drone detection and tracking in distorted infrared images

Abdelouahid Bentamou, Abbass Zein-Eddine, Oussama Messai, Yann Gavet

REAL-TIME DRONE DETECTION AND TRACKING IN DISTORTED INFRARED IMAGES

Abdelouahid Bentamou, Abbass Zein-Eddine*, Oussama Messai*, Yann Gavet**

*Mines Saint-Etienne, Univ Lyon, CNRS, UMR 5307 LGF, Centre SPIN, F - 42023 Saint-Etienne, France

ABSTRACT

With the increasing use of drones for various applications, their detection and tracking have become critical for ensuring safety and security. In this paper, we propose an algorithm for detecting and tracking drones from infrared (IR) images in challenging conditions such as noise and distortion. Our algorithm involves YOLOv7 for drone detection and utilizes the SORT algorithm for real-time tracking. To detect distortion in the drone images, we employed a vision transformer in parallel with a customized CNN. The experimental results demonstrate the effectiveness of our approach in challenging conditions and highlight the potential for future developments in drone detection and tracking using deep learning techniques. We achieve a precision of 94.2%, a recall of 92.64%, and a mean average precision (mAP) of 92.6% on the provided test data. *The implementation code can be found at: https://github.com/a-bentamou/Drone-detection-and-tracking*.

***Index Terms*—** Drone detection, drone tracking, distortion detection, YOLOv7, SORT, vision transformer

1. INTRODUCTION

Unmanned aerial vehicles (UAVs), or drones, have become increasingly popular over the years due to their versatility and affordability. Drones are used in various applications, including aerial photography, search and rescue missions, crop monitoring, and package delivery [1]. They have also been used in disaster management and relief efforts, as they can provide critical situational awareness and deliver essential supplies to hard-to-reach areas [2]. However, this technology has also given rise to new security challenges, particularly in the field of surveillance. Drone-based threats, such as espionage, smuggling, and terrorist attacks, have become a serious concern for governments and private organizations alike [3].

Different sensors have used to detect drones such as radar which faced difficulties due to the weak electromagnetic signals transmitted by drones, acoustic and radio frequency-based detection were also costly and inaccurate [4]. In addition, different deep learning-based systems for drone detection have been developed using various techniques. Some use two-stage detection algorithms such as Faster R-CNN [5], while others use one-stage detection algorithms like Single Shot Detector (SSD) [6] and You Only Look Once (YOLO) versions (v2 [7], v3 [8], v4 [9], v5 [10]). One-stage detectors are known for their speed and real-time operation. YOLO [11] has gained popularity in many computer vision tasks due to its fast and accurate detection, making it suitable for real-time implementation. Recently, YOLOv7 [12] has been developed, which significantly improves the algorithm's performance.

A solution to drone detection is the use of infrared (IR) imaging-based drone detection and tracking systems. These systems rely on the detection of heat signatures emitted by drones, which can be captured using IR cameras. However, drone detection in distorted surveillance videos poses several challenges, including the small size and speed of drones, their flight characteristics, and the limitations of traditional sensors. Environmental factors, such as background clutter and weather conditions, can also affect detection accuracy, as can noise and distortion in surveillance videos such as low light conditions and motion blur. Addressing these challenges requires innovative solutions that incorporate advanced sensor technologies, data processing algorithms, and machine learning techniques.

In this paper, we present an approach for drone detection and tracking using IR imaging in distorted surveillance videos. We propose to use YOLOv7 detector with Simple Online and Realtime Tracking (SORT) algorithm [13] which can effectively detect and track drones in complex scenarios, such as those with cluttered backgrounds and occlusions. Our approach incorporates a variety of techniques, including drone detection, object tracking, and noise distortion detection to achieve accurate and reliable detection and tracking of drones.

The remainder of this paper is organized as follows. In Section 2, we describe the proposed approach in detail. In Section 3, we present the experimental results and evaluate the performance of our approach. Finally, in Section 4, we conclude the paper and discuss future research directions.

2. PROPOSED APPROACH

In this section, we present the used dataset and the proposed methods for drone detection and tracking.

2.1. Dataset

The dataset used in this study is derived from the original "Drone detection dataset" [14]. It contains images extracted

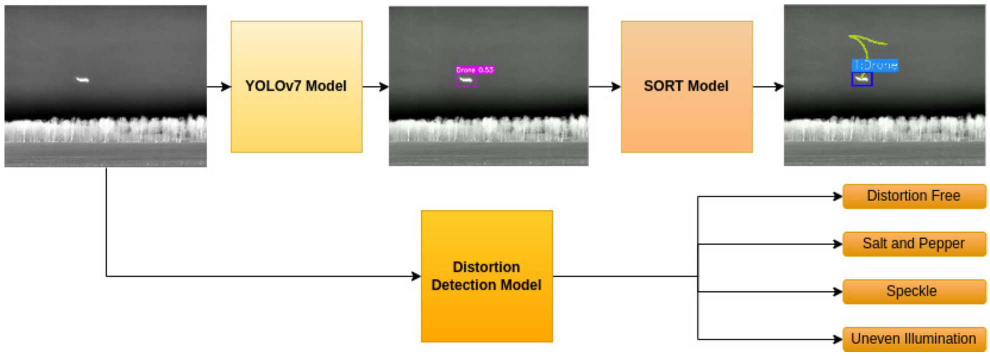

Fig. 1. Proposed neural network for drone detection and tracking.

from infrared (IR) videos at 30 frames per second and includes two classes: drones and birds. The images are of size 640×640 and are captured in challenging environments, such as thick forest cover, cloudy sky, fog, and mist. The dataset includes objects that are close or far-off from the field of view of the source camera as illustrated in Fig. 2.

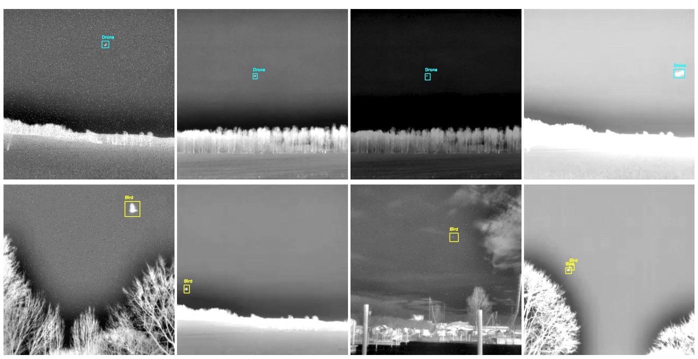

Fig. 2. Example of IR images from the proposed dataset. They illustrate some challenging unconstrained environments such as different backgrounds, forest cover, luminosity variations, and the various types of distortions. We can also see that objects (drone or bird) are close or far-off from the camera.

To enhance the algorithm's robustness, the dataset includes various types of distortions, including salt and pepper noise, speckle noise, and uneven illumination noise. The training data consists of 7908 images, of which 2748 are noised with salt and pepper noise, 2752 images with speckle noise, and 2752 images with uneven illumination noise. The training data includes 8179 bird bounding boxes and 484 drone bounding boxes. The number of images containing birds is significantly more than those containing drones, thus we are in the case of imbalanced data, which is a significant challenge in drone detection. To address this issue, we perform data augmentation during training process by adding new images including drones from the original dataset [14]. Finally, different type of distortions similar to that in the competition dataset are added to the new images.

Two levels of evaluation are performed in order to assess our proposed solution. First the level of drones detection. At this level we use the dataset provided by the competition and it comprises 726 images, containing 726 drone bounding boxes and no bird bounding boxes. The second level of evaluation is to assess the distortion detection method. Here we use the images provided by the competition in addition to other synthetic images that we create with various levels of distortion. For drone tracking task, there are no labelled videos in the provided dataset.

2.2. Proposed method

In this section, we describe the proposed method for infrared imaging-based drone detection and tracking in distorted surveillance videos. The proposed method involves three main steps as illustrated in Fig. 1: (1) drone detection using YOLOv7 algorithm, (2) drone tracking using SORT algorithm, and (3) detection of distortion in the IR images using a vision transformer and a customized CNN. Each step is explained in detail below.

2.2.1. Drone detection

The You Only Look Once (YOLO) algorithm [11] is a popular object detection method known for its high accuracy and real-time detection capabilities. Compared to other object detection algorithms like Region-Based Convolutional Neural Network (R-CNN) [15] and SSD [16], YOLO is faster and more precise. The latest versions of YOLO, such as YOLOv7 [12] and YOLOv8, achieve even higher average precision on the MS COCO dataset than the previous versions. In the context of detecting drones in surveillance videos, YOLOv7's real-time detection capability and high level of precision make it a suitable choice for this task. Its ability to accurately identify small, fast-moving objects like drones is particularly useful in surveillance scenarios where speed and accuracy are crucial for identifying potential security threats.

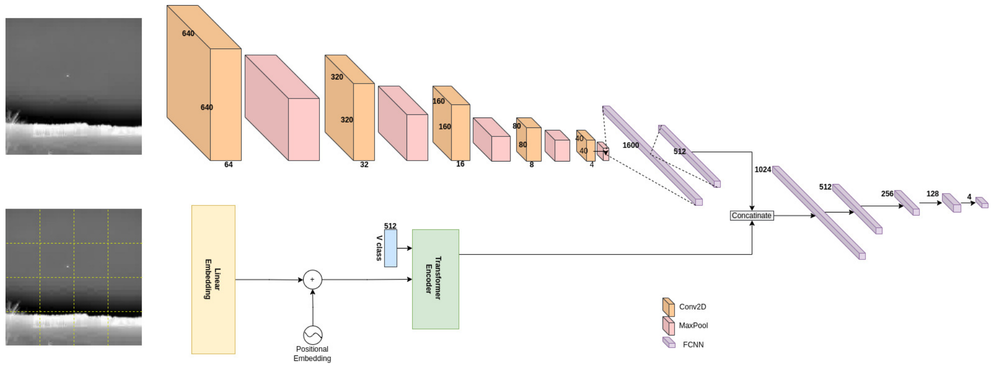

Fig. 3. Proposed neural network for distortion detection.

2.2.2. *Drone tracking*

The Simple Online and Realtime Tracking (SORT) algorithm [13] is a popular object tracking method that is designed to track objects in real-time using a single camera. SORT uses a combination of motion and appearance features to track objects and can handle occlusions, clutter, and other challenges associated with object tracking. SORT is a simple yet effective algorithm that can track multiple objects simultaneously and is widely used in various applications such as surveillance [17] and self-driving cars [18]. The algorithm is computationally efficient, making it suitable for real-time applications, and it has a high tracking accuracy and robustness, even in cluttered environments.

In drone tracking, SORT uses motion and appearance features to track the drone and predict its future position. The algorithm can handle the occlusion caused by objects in the drone's path and can track multiple drones simultaneously. SORT is also efficient at processing video data, allowing for real-time drone tracking in surveillance videos. Additionally, SORT can be combined with other algorithms, such as the YOLOv7 algorithm, to enhance drone detection and tracking accuracy. Overall, SORT algorithm provides an effective solution for drone tracking in surveillance videos, making it a valuable tool for security and surveillance applications.

2.2.3. *Noise distortion detection*

Deep learning methods distinguish between noise and image signals by learning the noise distribution features from a large number of noisy data samples. These approaches have been frequently employed for image dehazing, debluring, and so forth. In distortion detection problem, non-based deep learning [19] and deep learning based methods [20] have been proposed, with the latter showing superiority due to automated feature extraction and the availability of large datasets for training. More recently, Vision Transformer (ViT) [21] has grown in popularity due to its remarkable performance. It is now commonly used in image recognition tasks. ViT is a deep neural network architectural type. The ViT, has been inspired from natural language processing (NLP) transformer, which uses self-attention mechanisms to capture relationships between different parts of the input sequence. In this work, we develop a ViT that takes an image as input and divides it into patches, which are then fed through a sequence of transformer layers. These transformer layers process the patches and extract high-level features, which are then fed through a final classification layer to predict the image label. ViT has been shown to achieve state-of-the-art performance on several benchmark datasets, demonstrating its potential as a powerful tool for image recognition tasks.

In the context of IR images, ViT can be used in combination with customized Convolutional Neural Networks (CNNs) to detect noise and distortion in the images (c.f Fig. 3). High-level features, extracted from the IR image using ViT and CNN in parallel, are concatenated to feed fully connected layers in order to classify the image input into four classes (detection free, Salt and Pepper, Spackle and Uneven Illumination noises). The combination of these two techniques improve the accuracy for noise and distortion classification in IR images.

3. EXPERIMENTAL RESULTS

In this section, we present the results of our proposed algorithm. We evaluate the performance of the drone detection using various metrics such as mean Average Precision (mAP), precision, and recall. Additionally, we analyze the accuracy of the algorithm in detecting different types of distortions. We notice that this method was part of the ICIP 2023 challenge titled "Infrared Imaging-based Drone Detection and Tracking in Distorted Surveillance Videos". We are proud to report that our method achieved second place in this competition.

3.1. Drone detection

Table 1 presents the results in term of the precision, recall and mAP metrics on test data where the images are noised with different noise distortions. These evaluation results demonstrate the high accuracy of the proposed model in detecting drones under different noise distortions. The precision, recall, and mAP metrics indicate that the model performs consistently well, achieving a precision score of 94.2%, a recall score of 94.62%, and an mAP score of 92.6%. These results suggest that the proposed model is robust and effective in detecting drones in various scenarios and under different noise distortions. Moreover, the precision-recall curve plotted in Fig. 4 further supports the high accuracy of the model, as the curve tends towards the upper-right corner, indicating the model's high precision and recall. Overall, the evaluation results demonstrate that the proposed model can be a valuable tool for drone detection in real-world applications.

Table 1. The results of the evaluation for an intersection over union (IoU) = 0.5 on testing data (en %).

Class	Precision	Recall	mAP
Drone	94.2	94.62	92.6

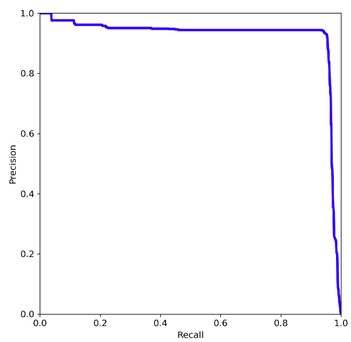

Fig. 4. Precision-recall curve of drone class on testing data.

3.2. Drone tracking

We note that the provided dataset lacks labelled drone tracking videos for algorithm evaluation. However, we apply our algorithm on videos from the original dataset [14] that do not contain any tracking labels. Fig. 5 illustrates some example frames of the algorithm's tracking performance on a video. As we can see from the figure, the proposed algorithm successfully tracks the drone in all video frames. The SORT algorithm is efficient and robust for drone tracking and it is able to track the drone accurately without the need a training phase or labelled videos.

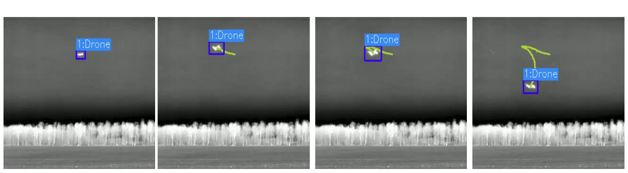

Fig. 5. Examples of drone tracking from different frames.

3.3. Noise distortion detection

In order to evaluate the performance of our noise distortion detection, we present the confusion matrix in Fig. 6. The confusion matrix shows the classification results of the test data with different types of noise distortions: speckle noise, salt and pepper noise, and uneven illumination noise. The diagonal elements in the matrix represent the percentage of correctly classified images, while the off-diagonal elements represent the percentage of misclassified images. From the confusion matrix, we can see that our proposed algorithm achieved high precision in classifying the distorted images. Specifically, 91% of speckle noise, 100% of salt and pepper noise, and 93% of uneven illumination noise were correctly classified. The algorithm achieved an accuracy of 92% indicating that our method is effective in detecting different types of noise distortions in IR images.

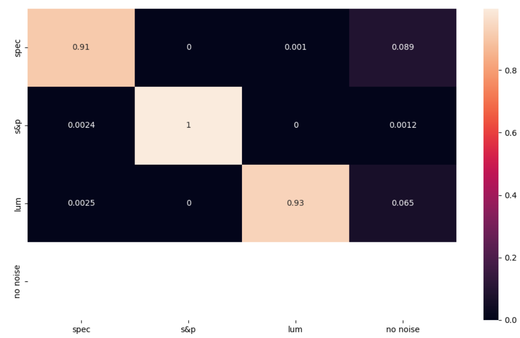

Fig. 6. Confusion matrix for noise distortion detection in IR images.

4. CONCLUSION

In this paper, we proposed an algorithm for drone detection from IR images using YOLOv7 and deep learning-based noise distortion detection using a vision transformer and a customized CNN. We also used data augmentation to handle the class imbalance problem, which improved the accuracy of our algorithm. Our experiments showed that our approach for drone detection achieved a precision, recall, and mAP of 94.2%, 94.62%, and 92.6%, respectively, on the provided test data. Moreover, the proposed noise distortion detection achieved a high classification accuracy of 91% for speckle noise, 100% for salt and pepper noise, and 93% for uneven illumination noise. Overall, our results demonstrate the effectiveness of our approach in drone detection and noise distortion detection, which can be useful in various applications, such as surveillance, search and rescue, and monitoring of critical infrastructure. Future work can involve exploring the use of other deep learning architectures and investigating the performance of our approach on different datasets and environmental conditions.

5. REFERENCES

[1] U. Seidaliyeva, M. Alduraibi, L. Ilipbayeva, and A. Almagambetov, “Detection of loaded and unloaded uav using deep neural network,” in *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, pp. 490–494, IEEE, 2020.

[2] C. Wankmüller, M. Kunovjanek, and S. Mayrgündter, “Drones in emergency response–evidence from cross-border, multi-disciplinary usability tests,” *International Journal of Disaster Risk Reduction*, vol. 65, p. 102567, 2021.

[3] J.-P. Yaacoub, H. Noura, O. Salman, and A. Chehab, “Security analysis of drones systems: Attacks, limitations, and recommendations,” *Internet of Things*, vol. 11, p. 100218, 2020.

[4] F.-L. Chiper, A. Martian, C. Vladeanu, I. Marghescu, R. Craciunescu, and O. Fratu, “Drone detection and defense systems: survey and a software-defined radio-based solution,” *Sensors*, vol. 22, no. 4, p. 1453, 2022.

[5] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma, and M. Blumenstein, “Drone detection in long-range surveillance videos,” in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, IEEE, 2019.

[6] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, and E. T. Matson, “Real-time and accurate drone detection in a video with a static background,” *Sensors*, vol. 20, no. 14, p. 3856, 2020.

[7] C. Aker and S. Kalkan, “Using deep networks for drone detection,” in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, IEEE, 2017.

[8] H. R. Alsanad, A. Z. Sadik, O. N. Ucan, M. Ilyas, and O. Bayat, “Yolo-v3 based real-time drone detection algorithm,” *Multimedia Tools and Applications*, vol. 81, no. 18, pp. 26185–26198, 2022.

[9] S. Singha and B. Aydin, “Automated drone detection using yolov4,” *Drones*, vol. 5, no. 3, p. 95, 2021.

[10] B. Aydin and S. Singha, “Drone detection using yolov5,” *Eng*, vol. 4, no. 1, pp. 416–433, 2023.

[11] 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*, pp. 779–788, 2016.

[12] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” *arXiv preprint arXiv:2207.02696*, 2022.

[13] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in *2016 IEEE international conference on image processing (ICIP)*, pp. 3464–3468, IEEE, 2016.

[14] F. Svanström, F. Alonso-Fernandez, and C. Englund, “A dataset for multi-sensor drone detection,” *Data in Brief*, vol. 39, p. 107521, 2021.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 21–37, Springer, 2016.

[17] I. Perera, S. Senavirathna, A. Jayarathne, S. Egodawela, R. Godaliyadda, P. Ekanayake, J. Wijayakulasooriya, V. Herath, and S. Sathyaprasad, “Vehicle tracking based on an improved deepsort algorithm and the yolov4 framework,” in *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pp. 305–309, IEEE, 2021.

[18] A. H. Abdel-Gawad, A. Khamis, L. A. Said, and A. G. Radwan, “Vulnerable road users detection and tracking using yolov4 and deep sort,” in *2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, pp. 140–145, IEEE, 2021.

[19] O. Messai, F. Hachouf, and Z. A. Seghir, “Automatic distortion type recognition for stereoscopic images,” in *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, pp. 1–4, IEEE, 2019.

[20] N. Ahn, B. Kang, and K.-A. Sohn, “Image distortion detection using convolutional neural network,” in *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 220–225, IEEE, 2017.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.