Identification of drone thermal signature by convolutional neural network

Chong Yu Quan[1], Ong Le Wei Edmond[2] and Sutthiphong Srigrarom[3]

***Abstract*— This paper presents the work on drone detection and identification using thermal infrared emission, which is primarily aimed towards night operation. Through both indoor and outdoor trials, the characteristics of the thermal signature emitted by a drone when captured by a drone detection system is examined, and their implications on a machine learning problem are studied. Thermal maps are processed through a YOLOv3 based CNN model to detect and generate a bounding box around the thermal signature of the drone. The presented approach also seeks to utilise the characteristics of drone motion for more effective drone detection through machine learning.**

I. Introduction

Recent progress in the development of unmanned aerial vehicles (UAVs) led to increasing situations in which drones, such as quadcopters, could potentially become a serious security threat and powerful tool for illegal activities. Therefore, counter-UAV systems are required in numerous applications to detect approaching drones as early as possible.

Drone detection by a thermal camera is primarily aimed towards night operation or locations with poor lighting. Hence, drone detection using thermal infrared emission is an exciting emerging field. Our prior work [1] focuses on drone detection in the day using normal visual (RGB) cameras. However, these methods become less effective when the night fall and visibility drops. This gives rise to the need for UAV detection using thermal infrared camera for night operation, to be presented in this paper. We anticipate that while RGB cameras with a high resolution enables drone detection up to further distances in the day, surveillance at night can be performed with a thermal infrared camera. Figure 1 illustrates the difference in imaging performance between a RGB camera and a thermal camera, in capturing a drone from the same distance.

II. Literature review

Gim et al [2] examined the effectiveness of various sensors (including thermal cameras) under different weather and environmental conditions in Singapore. It is a very relevant source as it is done specifically in Singapore context, where the study for this paper is conducted.

There is, however, very limited number of literature study in the area of drone detection through thermal infrared imaging.

(a) RGB Camera

(b) Thermal Camera

Fig. 1. Drone imaging

In 2017, Andrasvsi et al [3] specifically sought to examine the thermal signature of a drone. This study is restricted to understanding the effectiveness of drone thermal imaging methods, without any reference to drone detection through machine learning.

Mueller [4] used both visual and thermal cameras for drone detection in both day & night operations. However, he adopted a different machine learning approach; thereby presenting a different set of trade-offs.

Recently, in 2020, Svanstrom [5] presented his master

[1]Yu Quan Chong is with National University of Singapore, 5A Engineering Drive 1, Singapore 117411 e0321496@u.nus.edu

[2]Edmond Ong is with National University of Singapore, 5A Engineering Drive 1, Singapore 117411 e0273904@u.nus.edu

[3]Sutthiphong Srigrarom is with National University of Singapore, 5A Engineering Drive 1, Singapore 117411 spot.srigrarom@nus.edu.sg

thesis on Drone detection and classification using infrared imagery, through a YOLOv2 model. Nonetheless, this study is mainly focused on sensor fusion. Qi et al [6] presented their work on detecting drones from infrared images through a fast-saliency method and a histogram of oriented gradients (HOG) descriptor. Importantly, drone detection is done on each frame individually in the two above-mentioned sources, thereby neglecting any information pertaining to the motion of the drone over time.

III. Drone Thermal Imaging

A. Drone thermal signature

The thermal signature of a drone is highly dependent on its design characteristics and importantly, the quality of the thermal camera. The range for drone detection of the thermal camera used in this project is found to be around 6 meters in both indoor (air conditioned) and outdoor (night time) settings. This subsection seeks to establish some of the general characteristics of drone thermal imaging and their implications on a machine learning problem.

The thermal signature of a drone is found to vary across both time and distance from sensor, with the latter being more significant. At close range (2-3m), the infrared radiation emitted by the motors and the battery can be detected (see Figure 2). At further distances (>4m), the thermal signature detected is primarily due to the battery, giving rise to a fairly generic circular profile (see Figure 2). It is therefore important to validate the model's ability to not mistake the motor as a drone. While the heat signature of the drone is expected to change over time as it heats up, the time required for the drone to generate an adequate heat signature to be detected from the time it is turned on is fairly negligible. Other possible variants of the heat signature emitted by a drone is presented in Figure 3.

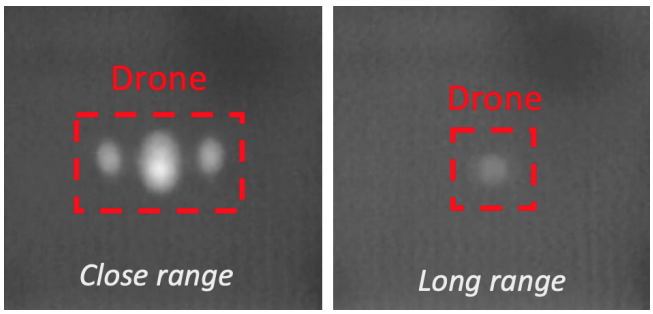


Fig. 2. Thermal signature of drone

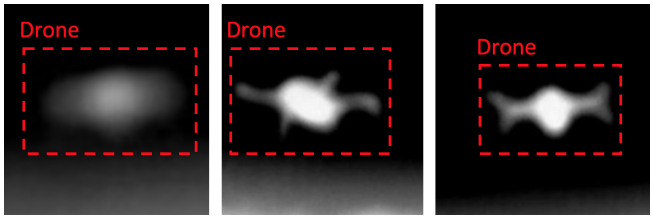


Fig. 3. Other possible variants of drone thermal signatures

Collectively, these point to the importance of considering the motion of the drone for the model to make accurate predictions, rather than solely relying on examining the heat signature emitted. The following outlines our approaches for drone detection and identification by thermal infrared camera:

1) Manually fly the aerial target (drones) in the laboratory (indoor) and outdoor at night at varying distances.
2) Initial detection of the aerial target (drones) by thermal infrared camera.
3) Store a thermal imaging video feed into a replay buffer real-time.
4) Extract a specified number of frames prior to the frame at the current time step from the video feed and concatenate them (in the channels dimension) to form a stacked image.
5) Implement a YOLOv3–based CNN model to obtain the bounding box for the drone from the stacked image.

The stated approach aims to utilise the correlated image features from the prior frames to encapsulate the nature of motion (e.g. velocity, path) of a thermal signature as an identifier of an object in motion. With a sufficiently large and diverse training data, including other potential objects (e.g. birds), the model should be able generalise across various unique nature of motion for identification purposes.

B. Methodology

In order to gauge the feasibility of the approach, test the hardware and collect data for supervised machine learning, an indoor trial in the laboratory was conducted first. The indoor environment was selected as the temperature of the air was generally low and constant given air-conditioning and with minimal thermal noise. The designated room with 6m by 6m space was divided into 1m by 1m grids, where the infrared camera was placed at the coordinates of $(3,0)$ as shown in Figure 4 below.

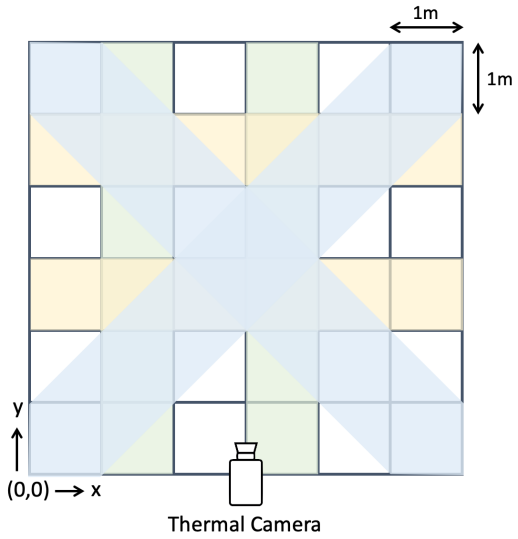


Fig. 4. Top view of flight test set up

To maximise the possible variety of motions of the drone captured in the data collection, the flight path of the drone was devised as follows:

1) Flight path of drone was restricted in the Y axis and unrestricted in the X and Z axis, allowing for roll, pitch, yaw and change in altitude, as illustrated by the yellow regions in Figure 4. This ensured sufficient data collected of the drone moving along a specific depth.
2) Flight path of drone was restricted in the X axis and unrestricted in the Y and Z axis, allowing for roll, pitch, yaw and change in altitude, as illustrated by the green regions in Figure 4. This ensured sufficient data collected of the drone moving in and out from varying depths.
3) Diagonal movements of the drone across grid unrestricted in the Z axis, allowing for roll, pitch, yaw and change in altitude as illustrated by the blue regions in Figure 4. This ensured sufficient data collected of with motion along all three axis.
4) Free-for-all flight path within the whole 6 x 6 grid to capture other forms of erratic motions (e.g. sudden and random acceleration).

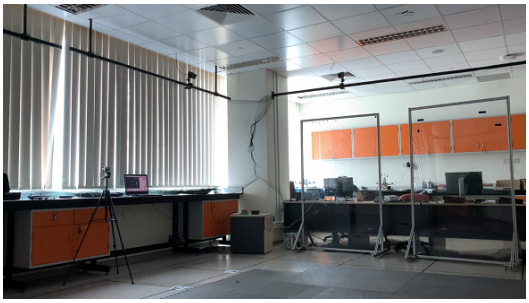
In addition, two sets of data were gathered in an outdoor environment at night: one over an open field and the other over a reservoir, as illustrated in Figure 5(a) and 5(b). These data were used as additional test sets that would be significantly more challenging that those from the indoor environment given the presence of background thermal noise from objects such as buildings and lampposts that are not found in the indoor environment.

C. Test setup

Thermal imaging was conducted using an Optris PI Lightweight thermal camera with Optris PIX Connect software, equipped with a wide-angle lens. Flight tests were conducted using a DJI Mavic Mini, powered by a 2400mAh 7.2V Li-ion 2S battery. The thermal map output was configured to operate within a fixed temperature range so that each temperature reading could be represented by a fixed color tone. This simplified the machine learning process.

(a) Reservoir

(b) Open Field



(c) Laboratory

Fig. 5. Test set-up

IV. Model Architecture

This paper presents an implementation of a convolutional neural network (CNN) for drone detection. YOLOv3 was selected in view of its well established track record in attaining a good mean average precision (mAP) score within a relatively shorter inference time [7], [8]. A short inference time is of critical importance in implementing real-time drone detection with low latency.

Importantly, the model is adapted to take into account the motion of the drone through an input of a concatenated (in the channels dimension) stack of frames in chronological order. An overview of the model architecture is presented in Figure 6.

Ideally, data should be gathered at a high frames per second (fps) to adequately encapsulate the drone motion into stepwise discrete frames. However, there exists a trade off during model training with regards to computational load and time. In addition, more labelling is required to generate the dataset for supervised learning. Therefore, a frame rate of 5fps was selected.

Latency is an important point of consideration in the implementation of a real-time drone detection system. It is a function of both the number of prior frames and the frame rate at which data is gathered. In this approach, a balance has to be made between the extent to which motion information is made available to the model for each prediction (per frame) and the resultant latency. In this model, the input consisted of a total of five frames (one frame for prediction and four previous frames). Nonetheless, it is important to note that the abovementioned latency applies only for the model's ability in distinguishing the thermal signatures of objects through motion. The implemented model still retains the ability of drone detection based on the conventional single image object detection, namely through the visual features

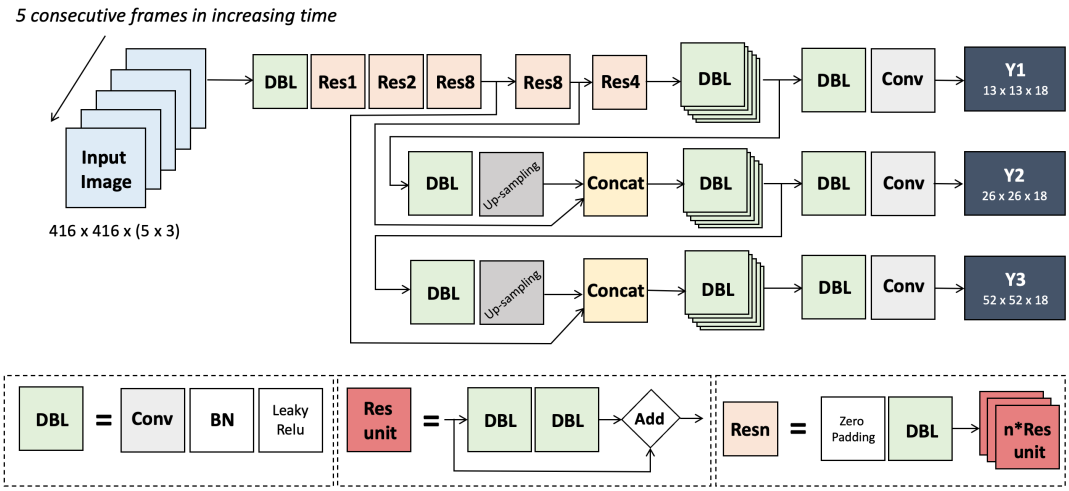


Fig. 6. YOLOv3 model architecture overview [7]

of the thermal signature. On an additional note, the problem of latency can be minimised by utilising a thermal camera with a greater reach, thereby reducing the chances of a drone going out of its field of view.

V. Model Training

A. Dataset

Data was gathered over a total flight duration of 1583 seconds, giving a total of 7913 frames. 68.8% of the data-set is used for model training while 8.4% and 22.8% of the data-set is used for validation and testing respectively. All the data gathered in this data-set was obtained in a controlled environment (air conditioned laboratory).

B. Losses

In this model, loss for the bounding boxes coordinates is obtained using generalised intersection over union (IoU) instead of the conventional IoU (otherwise referred to as Jaccard index). Conventional IoU presents an important disadvantage as the weights in the model are not modified in the event that the two bounding boxes (prediction and truth) do not intersect. Essentially, this represents a lost opportunity for the model to learn from the distance between these two bounding boxes. Generalised IoU accounts for this flaw by considering the smallest ellipsoid, C, enclosing the two bounding boxes, A and B, as shown in equation 2 [9].

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$GIoU = IoU - \frac{|C \setminus A \cup B|}{|C|} \tag{2}$$

In the context of bounding boxes, the following pseudocode shown in Figure 7 adapted from the work of Rezatofighi et al is used to generate GIoU losses for bounding boxes regression [9].

The losses for the objectiveness score and class probabilities are obtained using the sigmoid cross entropy loss

Algorithm 2: IoU and $GIoU$ as bounding box losses

input : Predicted B^p and ground truth B^g bounding box coordinates:
$B^p = (x_1^p, y_1^p, x_2^p, y_2^p), \quad B^g = (x_1^g, y_1^g, x_2^g, y_2^g)$.

output: $\mathcal{L}_{IoU}, \mathcal{L}_{GIoU}$.

For the predicted box B^p, ensuring $x_2^p > x_1^p$ and $y_2^p > y_1^p$:
$\hat{x}_1^p = \min(x_1^p, x_2^p), \quad \hat{x}_2^p = \max(x_1^p, x_2^p),$
$\hat{y}_1^p = \min(y_1^p, y_2^p), \quad \hat{y}_2^p = \max(y_1^p, y_2^p).$

Calculating area of B^g: $A^g = (x_2^g - x_1^g) \times (y_2^g - y_1^g)$.

Calculating area of B^p: $A^p = (\hat{x}_2^p - \hat{x}_1^p) \times (\hat{y}_2^p - \hat{y}_1^p)$.

Calculating intersection $\mathcal{I}$ between B^p and B^g:
$x_1^{\mathcal{I}} = \max(\hat{x}_1^p, x_1^g), \quad x_2^{\mathcal{I}} = \min(\hat{x}_2^p, x_2^g),$
$y_1^{\mathcal{I}} = \max(\hat{y}_1^p, y_1^g), \quad y_2^{\mathcal{I}} = \min(\hat{y}_2^p, y_2^g),$
$$\mathcal{I} = \begin{cases} (x_2^{\mathcal{I}} - x_1^{\mathcal{I}}) \times (y_2^{\mathcal{I}} - y_1^{\mathcal{I}}) & \text{if} \quad x_2^{\mathcal{I}} > x_1^{\mathcal{I}}, y_2^{\mathcal{I}} > y_1^{\mathcal{I}} \\ 0 & \text{otherwise.} \end{cases}$$

Finding the coordinate of smallest enclosing box B^c:
$x_1^c = \min(\hat{x}_1^p, x_1^g), \quad x_2^c = \max(\hat{x}_2^p, x_2^g),$
$y_1^c = \min(\hat{y}_1^p, y_1^g), \quad y_2^c = \max(\hat{y}_2^p, y_2^g).$

Calculating area of B^c: $A^c = (x_2^c - x_1^c) \times (y_2^c - y_1^c)$.

$IoU = \frac{\mathcal{I}}{\mathcal{U}}$, where $\mathcal{U} = A^p + A^g - \mathcal{I}$.

$GIoU = IoU - \frac{A^c - \mathcal{U}}{A^c}$.

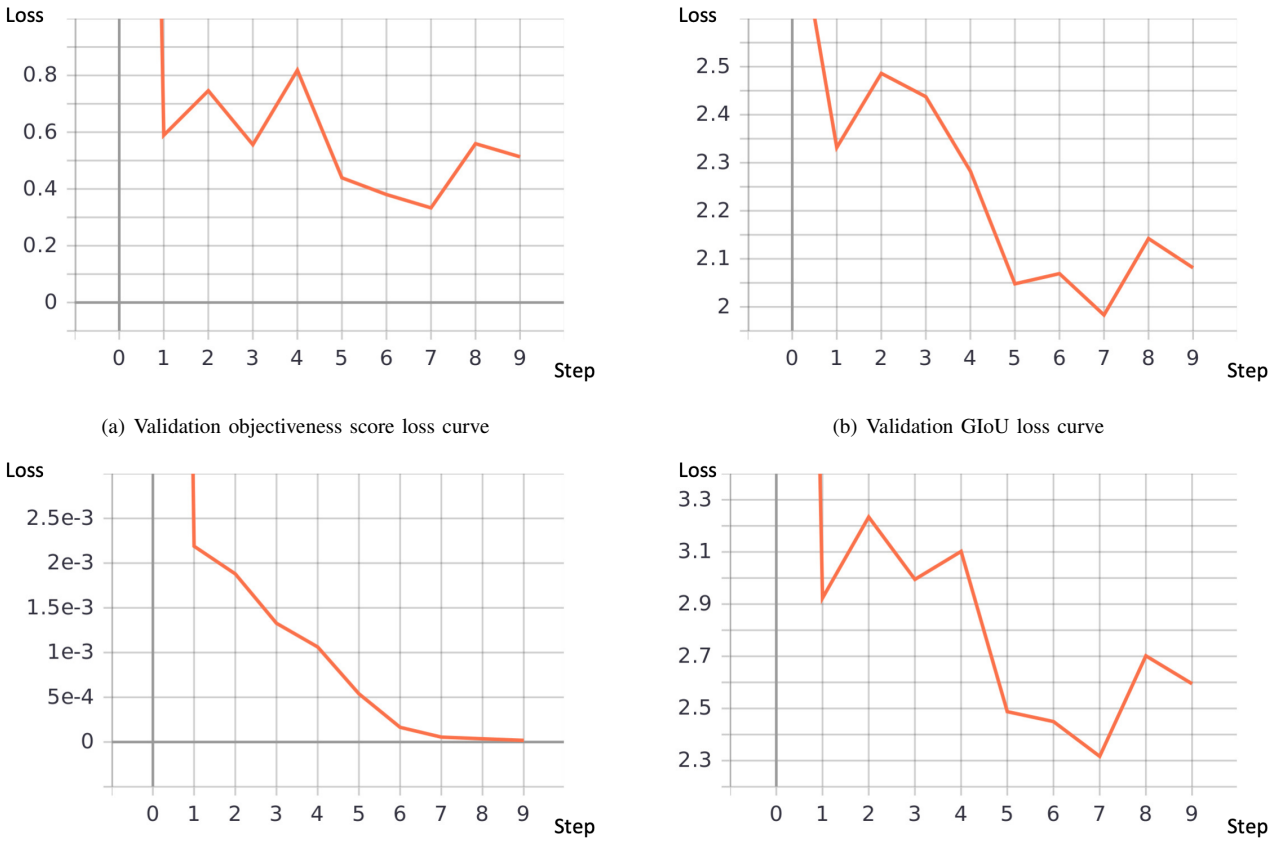
$\mathcal{L}_{IoU} = 1 - IoU, \quad \mathcal{L}_{GIoU} = 1 - GIoU.$

Fig. 7. Pseudocode for GIoU for bounding boxes [9]

as shown in equation 3, where y_i are the labels, $\hat{y}_i$ are the predictions and $S(x)$ is the sigmoid function defined in equation 4

$$-\frac{1}{m}\sum_{i=1}^{m} y_i log(S(\hat{y}_i)) - (1 - y_i) log(S(1 - \hat{y}_i)) \tag{3}$$

$$S(x) = \frac{1}{1 + e^{-x}} \tag{4}$$



(a) Validation objectiveness score loss curve

(b) Validation GIoU loss curve

(c) Validation probability loss curve

(d) Validation total loss curve

Fig. 8. Validation losses

C. Learning Rate

In the early training stages, a constant learning rate of 10^{-4} was initially utilised for training. However, such an approach led to the problem of exploding gradients for GIoU loss, where the updates for the model weights are untenably large, resulting in a diverging model. This is likely due to divisibility of very small values for C in computing the GIoU losses. This problem was solved through hyper-parameter tuning. This model applies a scheduled learning rate which decreases with a cosine annealing as presented in equation 5, where η^i_{max} and η^i_{min} are the ranges of the learning rate and T_{cur} is the current epoch [10].

$$\eta_t = \eta^i_{min} + \frac{1}{2}(\eta^i_{max} - \eta^i_{min})(1 + cos(\frac{T_{cur}}{T_i}\pi)) \quad (5)$$

The learning rate curve for this model is presented in Figure 10. A decreasing learning rate was utilised to optimise learning efficiency, since the losses and corresponding model weight modifications are expected to be greatest at the start and decreasing over time as the model was being trained.

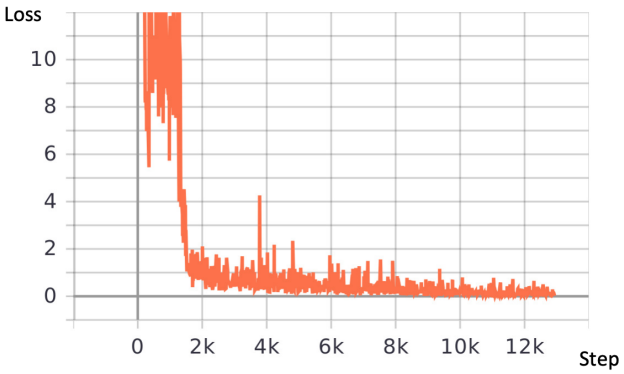
Figure 8 and 9 illustrate the training and validation losses for the model. From training, the order of magnitude of the objectiveness score was found to be the largest contributor to the total losses during early stages of training as a the majority of the bounding boxes proposed do not contain any objects of interest, since the only object is the drone. Nevertheless, with the help of a scheduled learning rate, the model is able to converge quickly within 10 epochs of training while avoiding the exploding gradients problem. The best performing model based on the validation loss, which is the model after epoch 7 as shown in Figure 8(d), is saved for testing.

VI. Results

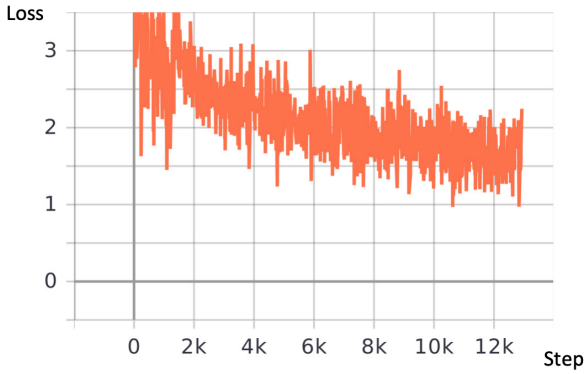
This model achieved a 98.3% success rate in generating a bounding box reasonable to a human interpreter. A precision of 1 (i.e. with no false positives) and a recall of 0.99 was attained, based on a 0.5 IoU threshold. On top of two frames where the model failed to identify the presence of a drone, the rest of the error in detection is attributed to wrongly drawn bounding boxes which encapsulates the drone partially, majority of which occurs at the edge of the frame as shown in Figure 11.

It is worth noting that the model identified the drone correctly without mistaking the heat signature of the motors as that of a drone in all instances (see Figure 12). In addition, the model is capable of correctly identifying the drone to a large extent when its thermal signature is varying as a result of a yawing motion (see Figure 13).

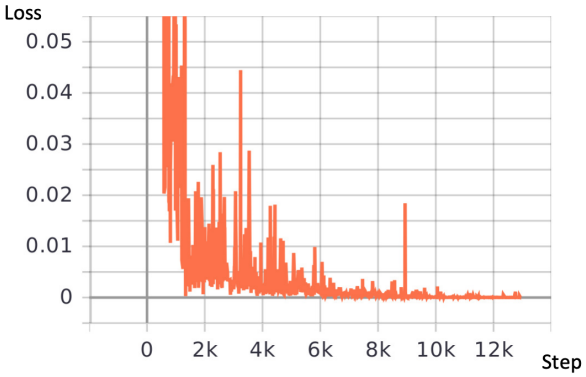
In furtherance to this testing, the model was tested on two data-sets (outside of the stipulated data-set) obtained



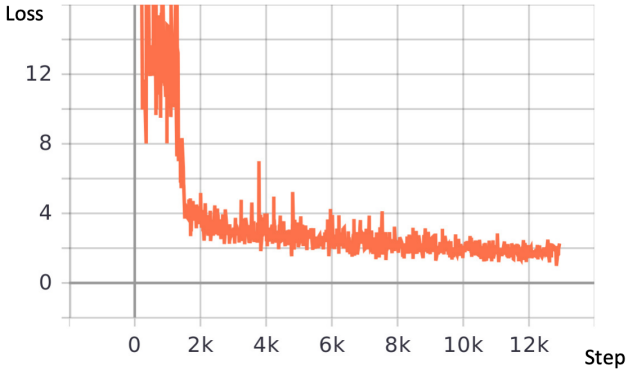
(a) Training objectiveness score loss curve



(b) Training GIoU loss curve



(c) Training probability loss curve



(d) Training total loss curve

Fig. 9. Training losses

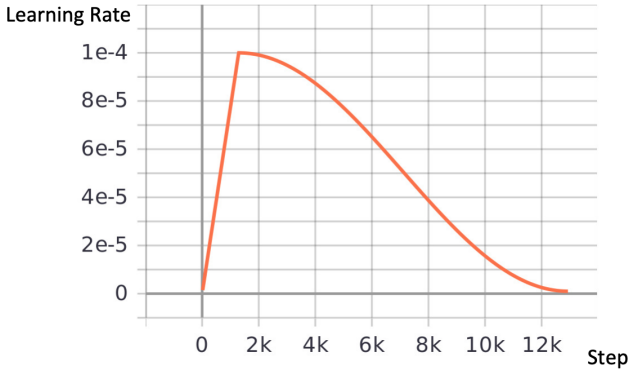


Fig. 10. Learning rate curve

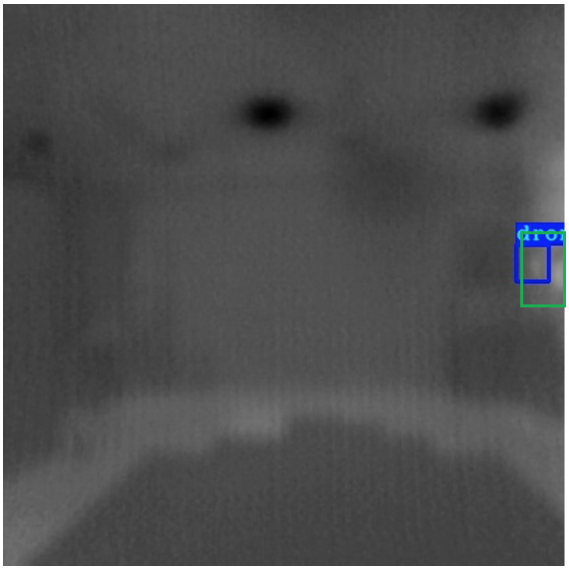


Fig. 11. Example of wrongly drawn bounding box, with correct bounding box in green

under outdoor nighttime condition. One data set was obtained with the drone flying over an open field while the other was obtained with the drone flying over a reservoir. As expected, the performance of the model in these two tests was significantly poorer as new conditions (which are not found in the training data-set) are introduced. The precision (P) and recall (R) values yielded for these two tests are $P = 0.8$ and $R = 0.5$ (over land) ,and, $P = 0.8$ and $R = 0.7$ (over water).

The different temperature profile of the outdoor environment is found to have introduced some false positives in the model as shown in Figure 14. In addition, the model's ability in distinguishing between a drone and its motors is found to be lowered, as shown in Figure 15.

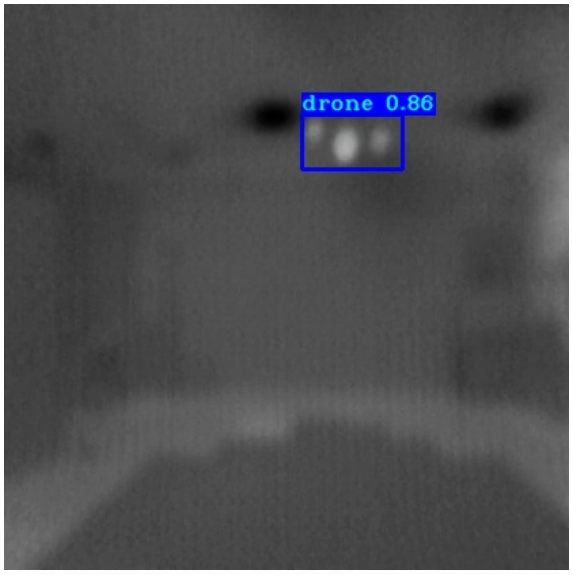


Fig. 12. Correct identification of drone

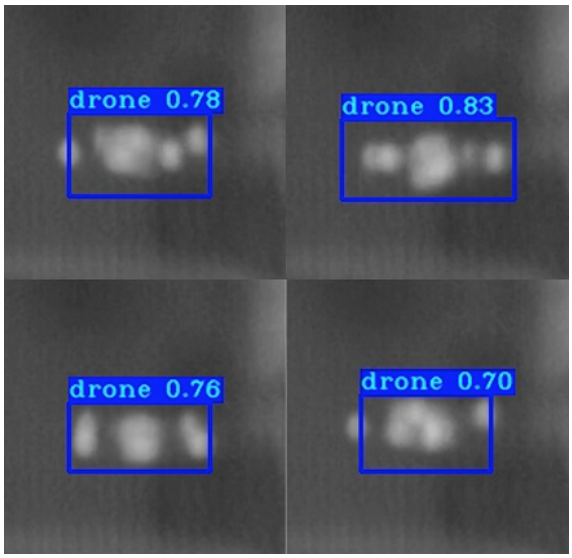


Fig. 13. Correct identification of drone in yawing motion

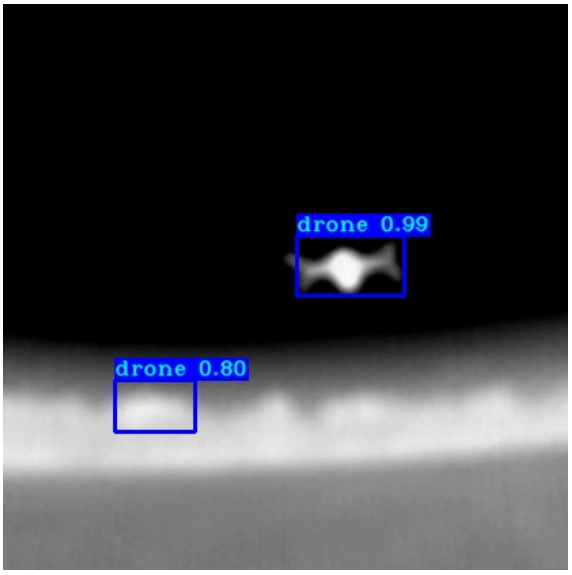


Fig. 14. False positive in drone detection

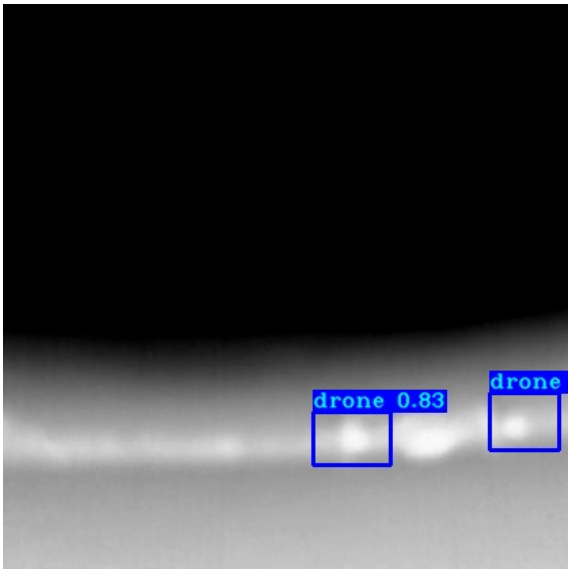


Fig. 15. False positive in drone detection (motors)

Most of the false negatives generated by the model occurs when the drone appears smaller (from a greater distance) as shown in Figure 16. This drawback is likely due to the lack of training data for such scenarios as the laboratory is limited in space.

Lastly, it is worth noting that the model successfully identified a drone which would not have been apparent to even a human interpreter without having been able to observe the previous frames to observe the subtle changes in the thermal map arising from drone motion. As shown in Figure 17, the elongated temperature profile of the drone identified happens to blend in with the environment. This observation strongly suggests that the ability to consider the motion of the drone through concatenating a number of frames in chronological order is present in the model.

VII. Discussion

The model's ability to effectively identify the thermal signature of a drone is validated to a large extent based on the high precision and recall values.

Despite the deterioration of the model's performance in the outdoor environment due to various factors stated in the previous section (e.g. the presence of other thermal noise), the loss in performance is within expectation given that the model is trained using a dataset that is obtained in a controlled indoor environment, which is vastly different from the thermal profile of the outdoor environment. The solution to improve the model's outdoor performance is trivially to obtain a significant data set that is representative of the actual outdoor environment where the drone detection is implemented. Ideally, by training and exposing the model

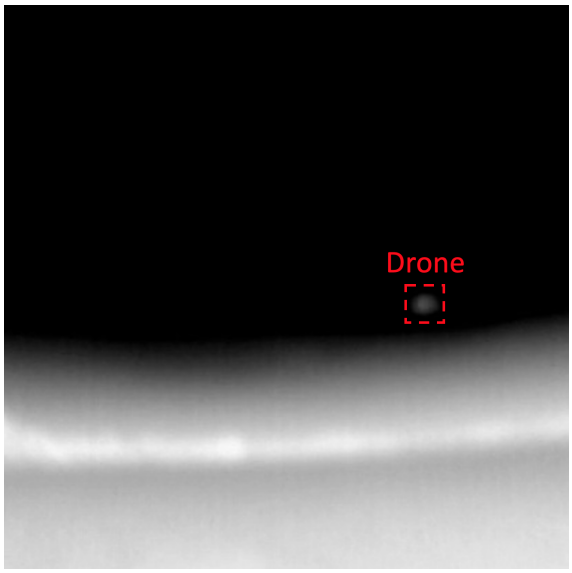


Fig. 16. False negative in drone detection

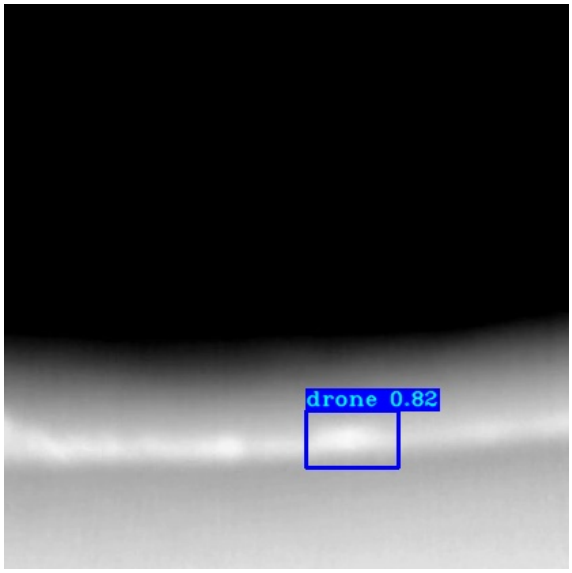


Fig. 17. Accurate drone detection based on motion

to a dataset with a good distribution of thermal profiles representative of the various outdoor environments that the model is expected to encounter, the model should generalise its drone identification abilities to the general outdoor environment.

Furthermore, given that the current dataset only contains one drone at any given frame, future work should expand and include datasets with a number of drones for training and validation of the model's robustness beyond a single drone. Following that, given that the current dataset also only includes one object class, namely the drone, future work should also expand the dataset to include other classes (e.g. birds, humans) in order to test and validate the model's ability to differentiate various thermal signatures in the outdoor environment. In particular, given that the thermal signature of that of a bird and a drone is largely similar visually at a distance (that of a circular thermal signature), the validation of the model's ability to differentiate between the stated two objects would strengthen the robustness of the concept of using stacked image frames to encapsulate motion as basis for differentiating objects with similar thermal signatures.

Lastly, it is important to recognise that the work done here is a proof of concept with regards to the use thermal imaging for drone detection. The setup with regards to the thermal camera used is not the most optimal camera for drone detection purposes given its detection of range of 6 meters. However, there are better thermal camera options that is unavailable for this research, such as the Optris PI 640 thermal camera with a potential instantaneous field of view (IFOV) of $10.2mm$ for a distance of $25m$ with the correct setup, hence significantly extending the potential detection range concept.

VIII. Conclusion

The proposed drone detection and identification based on machine learning using YOLOv3 based CNN model presented in this paper successfully identified the thermal signature of a drone with high precision and recall. In addition, preliminary observations strongly suggest that the approach of concatenating a number of frames in chronological order provided the model with meaningful information relating to the motion of a drone, giving rise to better drone identification. It is well within expectations that more training data is required for better performance under more diverse conditions.

References

[1] S. Srigrarom, S. M. Lee, M. Lee, F. Shaohui, and P. Ratsamee, "An integrated vision-based detection-tracking-estimation system for dynamic localization of small aerial vehicles," in *2020 5th International Conference on Control and Robotics Engineering (ICCRE)*. IEEE, 2020, pp. 152–158.

[2] L. Gim, E. K. Tiong, and H. Y. Elizabeth, "Performance challenges for high resolution imaging sensors for surveillance in tropical environment," *DSTA Horiz*, pp. 80–88, 2015.

[3] P. Andraši, T. Radišić, M. Muštra, and J. Ivošević, "Night-time detection of uavs using thermal infrared camera," *Transportation Research Procedia*, vol. 28, pp. 183–190, 2017.

[4] T. Müller, "Robust drone detection for day/night counter-uav with static vis and swir cameras," in *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VIII*, vol. 10190. International Society for Optics and Photonics, 2017, p. 1019018.

[5] F. Svanström, "Drone detection and classification using machine learning and sensor fusion," 2020.

[6] S. Qi, W. Zhang, and G. Xu, "Detecting consumer drones from static infrared images by fast-saliency and hog descriptor," in *Proceedings of the 4th International Conference on Communication and Information Processing*, 2018, pp. 62–66.

[7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[9] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.

[10] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.